MASTER EN SCIENCES INFORMATIQUES

Analyse des techniques de protection de la vie privée et de lutte contre le profilage sur Internet

Lafontaine, Arnaud

Award date:
2013

Awarding institution:
Universite de Namur

Facultés Universitaires Notre-Dame de la Paix, Namur
Faculté d'Informatique
Année académique 2012–2013

Analyse des techniques de protection de la vie privée et de lutte contre le profilage sur Internet

Arnaud Lafontaine

Promoteur : _____ (Signature pour approbation du dépôt - REE art. 40)
Pr. Claire Lobet-Maris

Mémoire présenté en vue de l'obtention du grade de
Master en Sciences Informatiques.

Remerciements

Tout d'abord, je tiens à remercier ma promotrice, le Pr. Claire Lobet-Maris, pour la proposition de ce mémoire, son enthousiasme et ses conseils avisés prodigués tout au long de celui-ci.

Je remercie également les membres de ma famille, mes amis et collègues pour leurs encouragements. Une pensée particulière est adressée à Nathalie, ma compagne, qui m'a toujours motivé et soutenu dans les meilleurs moments comme dans les plus durs.

J'aimerais une fois de plus exprimer ma reconnaissance à Nathalie Merlin mais aussi à Nicolas Goffaux, qui ont pris de leur temps afin de relire le présent document.

Enfin, je tiens à remercier tous ceux qui prendront le temps de s'intéresser à ce sujet captivant, que vous soyez expert en informatique ou simplement passionné.

Résumé

De nos jours, l'informatique et plus particulièrement Internet font partie du quotidien des Belges, s'immisçant dans leurs foyers et touchant leur intimité, quel que soit leur âge. Force est de constater que la vie privée y est régulièrement bafouée, entre les sociétés publicitaires avides d'informations personnelles permettant de cibler leur clientèle, les organismes gouvernementaux de tous continents souhaitant contrôler les moindres faits et gestes des internautes ou encore les fuites de données des entreprises à qui elles furent confiées.

Le premier chapitre de ce document a pour objectif de sensibiliser le lecteur à la nécessité de protéger sa vie privée sur Internet, de tenter de définir ce concept et d'en expliquer l'enjeu. La législation en vigueur dans ce domaine est abordée du point de vue du particulier mais également du gestionnaire d'un site Internet ou de tout autre service collectant des informations personnelles. Sera ensuite brièvement expliquée la manière dont les données collectées sont regroupées sous forme de profils afin de catégoriser les internautes.

Viennent ensuite, au second chapitre, les descriptions successives de différentes techniques qu'il est possible d'exploiter afin de collecter des informations relatives à un internaute et aux activités qu'il mène sur les différents sites qu'il visite. L'importance relative de chacune de ces techniques, sa capacité à fournir des informations, est illustrée dans un tableau comparatif, où il est observable que les données personnelles, la localisation géographique et les centres d'intérêt sont les catégories d'informations obtenues par le plus grand nombre de techniques.

Cette analyse est complétée par la proposition de diverses solutions de protection des informations dont la sécurité est mise en péril et qui ont été énoncées au chapitre précédent. Seuls les moyens de défense abordables et aisés à mettre en place sont exposés. Chacun d'entre eux est évalué selon divers critères, qui donneront lieu à une cotation, permettant de les comparer. Cette analyse révèle que les techniques les plus simples à mettre en place et les moins coûteuses sont parfois aussi les plus efficaces, comme illustré par le réglage de la configuration du navigateur Internet et l'installation d'extensions adéquates.

En conclusion, il est possible, pour qui le souhaite, de protéger efficacement sa vie privée à moindre efforts et à faible coûts, et souvent sans influencer en aucune manière la qualité de l'expérience de navigation.

Abstract

Nowadays, informatics, and especially Internet, have become part of people's every day life, whatever their age is, intruding their home and their intimate life. One has to admit that privacy is often violated by companies that are keen on getting personal information in order to target the right customers, by governemental organisations around the world willing to track and control the Internet user's acts and facts, or by enterprises, leaking confidential data they are in charge of.

The first chapter of this thesis aims at making the reader aware of the necessity to protect his Internet privacy and attempts to define the "privacy" concept and explain its importance. The legislation in force in this domain is approached, both from the point of view of the private individual, and the manager of an Internet site or any other service collecting private data. This is followed by a brief explanation on the way private data can be organized in profiles to categorize the Internet users.

In the second chapter, different techniques are described on how to retrieve and exploit information about an Internet user and his activities on the various web sites he is visiting. The relative importance of each of these techniques and their capacity to deliver information is illustrated in a comparative table, from which it can be observed that personal data, interests and geographical location are the information categories that are most often obtained by these techniques.

This analysis is completed in the third chapter by the proposal of various solutions to protect the earlier mentioned information security which is put at risk. The measures proposed are meant to be affordable and easily put in place. Each of them is evaluated according to different criteria, giving rise to a quotation and allowing to compare them against each other. This evaluation reveals that the simplest techniques in terms of cost and installation are sometimes the most efficient ones, as illustrated by the Internet browser configuration settings and the installation of adequate plug-ins.

As a conclusion, it is illustrated in this thesis that it is definitely possible to protect efficiently an Internet user's privacy with minimal efforts and against limited cost, and this often without impacting the Internet navigation experience and quality.

Table des matières

Glossaire

Acronymes

CPVP	Commission de la protection de la vie privée
CGI	*Common Gateway Interface*
DNS	*Domain Name System*
EFF	*Electronic Frontier Fondation*
FAI	Fournisseur d'accès à Internet
GPG	*GNU Privacy Guard*
HTTP	*HyperText Transfer Protocol*
HTTPS	*HyperText Transfer Protocol Secure*
IANA	*Internet Assigned Numbers Authority*
IETF	*Internet Engineering Task Force*
IP	*Internet Protocol*
ISP	*Internet Service Provider*
LCEN	Loi pour la Confiance dans l'Economie Numérique
LIR	*Local Internet Registry*
LSO	*Local Shared Object*
NIR	*National Internet Registry*
P3P	*Platform for Privacy Preference*
PET	*Privacy-Enhancing Technologies*
PGP	*Pretty Good Privacy*
RIR	*Regional Internet Registry*
SSL	*Secure Socket Layer*
TCP	*Transmission Control Protocol*
TOR	*Virtual Private Network*
UDP	*User Datagram Protocol*
VPN	*Virtual Private Network*

Introduction

A l'heure actuelle, alors que l'informatique et plus précisément Internet occupent une place sans cesse croissante dans notre quotidien, il est important de prendre conscience de l'importance de la maîtrise de ces technologies. Plus qu'un simple outil de communication passif, Internet est une véritable plateforme vivante où chaque information est analysée dans des buts aussi nombreux que variés. Les évolutions technologiques ayant implanté Internet au sein de tous les foyers, son accès s'effectue de nos jours dès le plus jeune âge, d'où l'importance de conscientiser ses utilisateurs, parents et enfants.

La nature même des outils permettant d'accéder à Internet ne laisse en rien présager des dangers inhérents à leur utilisation. Seul devant son ordinateur ou son *smartphone*, il n'est pas évident, pour l'utilisateur, que de nombreux acteurs sont impliqués dans toute action qu'il entreprend.

Une fois cette prise de conscience effectuée, le lecteur se posera probablement des questions relatives à la nature des données collectées à son sujet, de la raison pour laquelle cette collecte est effectuée, de l'utilité de telles informations pour une tierce personne ou encore de ses droits au regard de la protection de sa vie privée. Ces questions seront abordées tout au long du présent document en commençant par une mise en contexte factuelle illustrant l'ampleur que peuvent prendre les fuites d'informations et les techniques de surveillance, lorsqu'elles ne sont pas contrôlées.

De manière plus concrète, nous fournirons un aperçu des principales techniques usitées au quotidien afin de glaner des informations à propos des internautes, ce que celles-ci permettent de révéler et la manière dont elles peuvent être exploitées afin d'établir un profil relatif à l'internaute. Profil anonyme ou l'identifiant personnellement, profil vague ou très détaillé, de nombreuses informations fournies volontairement ou non par l'utilisateur permettent d'étoffer toujours plus les connaissances à son sujet.

La connaissance du visiteur permet aux entreprises présentes sur Internet d'établir des catégories types de clientèle. Cette catégorisation, due au profilage, bien que présentant au premier abord des avantages de simplicité pour l'internaute, a également un effet néfaste sur l'objectivité des services fournis, l'égalité de tout individu face à l'offre, et finalement donne une perception négative de l'outil qu'est Internet.

Il incombe à chaque individu de mettre en place personnellement sa couverture sur Internet et, pour y parvenir, le devoir de s'informer. Toute personne utilisant le réseau mondial est concernée par les tentatives constantes d'atteinte à la vie privée, qu'il s'agisse d'un utilisateur occasionnel ou au contraire d'un habitué des outils de télécommunication. C'est pourquoi nous établirons une liste non exhaustive des principaux outils et techniques accessibles à tous et permettant de mettre en place une protection efficace de la vie privée d'un internaute. Chaque outil sera tour à tour décrit, analysé et noté en fonction de ses capacités de protection.

Nombre des thèmes considérés tout au long de ce document abordent des domaines dont la compréhension et l'analyse complète peuvent à elles seules faire l'objet d'un mémoire, qu'il s'agisse de l'analyse éthique, juridique ou sociale de l'espionnage sur Internet, des techniques de profilage pouvant être mises en place ainsi que leurs mécanismes internes de fonctionnement, ou encore les nombreux outils de collecte ou de protection d'informations. Il faut garder à l'esprit que cet ouvrage n'a pas la prétention d'aborder tous les aspects et les composantes en jeu, et qu'il a pour vocation une sensibilisation du lecteur à la protection de sa vie privée sur Internet.

Chapitre 1

Mise en contexte

Les avancées technologiques dans les domaines de l'information et des télécommunications ainsi que la mondialisation et la vulgarisation d'Internet ont exacerbé un problème majeur : l'exposition des données privées de chaque internaute et la perte de contrôle de ces informations ou de leur exploitation.

Afin d'établir des bases concrètes et de cerner la problématique, quelques faits marquants de l'actualité seront présentés, sélectionnés pour leur caractère récent ou pour la polémique qu'ils ont suscité lors de leur divulgation.

Sera ensuite proposée une ébauche de ce que représente la vie privée, sa frontière avec les informations publiques et les enjeux personnels et sociaux qui en dépendent.

Finalement, le cadre juridique sera abordé, situant les individus et les organisations au sein de leurs droits et leurs devoirs respectifs l'un envers l'autre.

1.1 Faits d'actualité

Cette section répertorie quelques faits d'actualité, plus ou moins récents, en Belgique et dans le monde, qui illustrent la croissance constante de l'espionnage informatique, dûe en grande partie aux innovations technologiques. Ces faits permettront d'aborder ce document avec un intérêt d'autant plus important qu'ils concernent directement les internautes, quelqu'ils soient.

1.1.1 Le projet PRISM

En juin 2013, Edward Snowden, un informaticien consultant employé par les renseignements américain, rend publique, par l'intermédiaire des médias, des informations appartenant à la NSA (*National Security Agency* ou Agence de Sécurité Nationale américaine) et classées top-secrètes. Celles-ci concernent le programme de surveillance électronique américain **PRISM** ainsi que la collecte des métadonnées [1] des appels téléphoniques aux États-Unis.

Documents à l'appui, les quotidiens américains "Washington Post" et "The Guardian" révèlent alors que la NSA et le FBI (*Federal Bureau of Investigation* ou Bureau Fédéral d'Enquête) ont accès aux bases de données de grandes entreprises très influentes sur Internet, parmi lesquelles Facebook, Google, YouTube, Paltalk, Yahoo !, Microsoft, Skype, AOL et Apple, afin d'analyser les données de leurs utilisateurs sans nécessiter de mandat judiciaire et cela à partir du moment où la personne concernée ou l'un des participants à une communication n'est pas sur le territoire

1. Une métadonnée est une donnée servant à définir ou décrire une autre donnée (date et heure de création ou enregistrement de fichier, géolocalisation de l'endroit de création, nom de l'auteur d'un document, etc.)

américain [1]. Des millions de courriers électroniques, vidéos, photos et autres documents furent ainsi scannés par ce programme.

Les journaux rapportent également que l'opérateur téléphonique américain "Verizon" a été contraint de livrer à la NSA des métadonnées à propos des conversations de ses clients, comme les numéros d'appel ou les durées des communications [2].

Suite à ces révélations, qui ont produit une véritable onde de choc à l'échelle mondiale, le gouvernement américain a confirmé l'existence du programme PRISM, mis en place depuis 2007, et permettant aux deux agences de consulter les données des clients des entreprises précédemment citées. Contrairement à son prédécesseur, le **Terrorist Surveillance Program** mis en place après les attentats du 11 septembre 2001, le projet PRISM a obtenu l'accord d'opérer de la FISC [2] et est supervisé par celle-ci, conformément au FISA [3].

Présenté comme un outil de lutte anti-terroriste, les responsables se défendent et expliquent que "Les données sont stockées sur les serveurs de la NSA mais ne seront utilisées que s'il existe des soupçons précis.".

Plusieurs compagnies concernées ont répondu qu'elles ne fournissaient pas d'informations en vrac à la NSA, mais que chaque demande de renseignement devait cibler certains individus et être en accord avec le FISA.

Peu de temps après, en juillet 2013, "The Guardian" publia un article détaillé concernant le logiciel d'espionnage de données **XKeyscore** créé par la NSA [3]. Les moindres faits et gestes des internautes peuvent être collectés en temps réel grâce à plus de 700 serveurs localisés dans plusieurs dizaines de pays. Courriers électroniques, discussions instantanées, connexions aux réseaux sociaux, téléchargements ou encore de simples requêtes sur un moteur de recherche, peuvent être captés par ce logiciel ultra performant qui permet ensuite de remonter jusqu'à l'identité de l'internaute. Les données captées seraient conservées seulement entre trois et cinq jours, les informations suspectes étant quant à elles conservées plus longtemps. Le logiciel aurait permis de capturer plus de 300 terroristes depuis 2008.

Enfin, les documents dévoilés par Snowden révèlent également l'existence d'un autre mécanisme, allant de pair avec PRISM et nommé **Upstream**. Selon le "Washington Post", celui-ci permet à la NSA de collecter les communications et autres flux de données en temps réel via des équipements spécifiques permettant d'écouter et de puiser à même la fibre optique et les infrastructures réseau [4].

Attentats évités grâce à la divulgation de messages électroniques

Le projet PRISM aurait permis aux autorités belges d'arrêter, en décembre 2008, Malika El Aroud et Moez Garsallaoui, recrutés par l'association terroriste Al-Qaeda en 2007, et en pleine préparation d'un attentat-suicide dans un appartement de Bruxelles. En effet, selon CNN, ces deux arrestations sont notamment dues à l'échange de messages électroniques ainsi que plusieurs conversations téléphoniques interceptés par les autorités américaines et dévoilés ensuite aux autorités belges [5].

2. *Foreign Intelligence Surveillance Court* : tribunal composé de 11 juges désignés par le président de la Cour Suprême, qui est lui-même nommé par le président des Etats-Unis et dont les décisions sont classées secrètes.

3. *Foreign Intelligence Surveillance Act* : loi du Congrès des États-Unis d'Amérique de 1978 décrivant les procédures des surveillances physiques et électronique, ainsi que la collecte d'informations sur des puissances étrangères soit directement, soit par l'échange d'informations avec d'autres puissances étrangères.

1.1.2 Proposition de loi belge relative à l'enregistrement des courriers électroniques

Pressé par la Commission européenne, le gouvernement belge déposa au Parlement, en juillet 2013, un projet de loi imposant aux fournisseurs de télécommunications de stocker, pour une durée de un an, toutes les traces de communication transitant par leurs serveurs [6].

Il s'agit d'une transposition dans le droit belge de la directive 2006/24/CE du Parlement et du Conseil de l'Union Européenne du 15 mars 2006, imposant aux opérateurs de télécommunications et aux fournisseurs d'accès à Internet de conserver toutes les données de communication des citoyens afin de lutter contre le terrorisme et la grande criminalité.

Les échanges téléphoniques et par SMS, déjà sujets à ce type de suivi, voyaient conservées les données relatives aux coordonnées de la source et du destinataire, la durée des conversations ainsi que la date, l'heure et le lieu où les appels ont été passés ou les messages envoyés, mais pas le contenu du message. Ces données sont déjà gardées en mémoire par les opérateurs actuellement, notamment pour établir les facturations [7]. Nouveauté pour les échanges de courriers électroniques, les adresses IP d'où partent et arrivent les messages électroniques doivent dorénavant également être enregistrées.

La directive européenne a soulevée une polémique auprès des organisations professionnelles de médecins, avocats et journalistes, car pouvant porter préjudice au secret professionnel, ainsi qu'auprès de la Ligue des droits de l'homme qui y voit une atteinte à la présomption d'innocence.

1.1.3 Les Google cars

En mai 2007, Google lançait le service "Google Street View" afin de compléter ses célèbres "Google Maps" et "Google Earth". Cette nouvelle application est dite de *Mobile Mapping*, technologie par laquelle un véhicule équipé de caméras et/ou de scanners peut enregistrer numériquement toutes les données d'une route spécifique, notamment par la prise de photos à 360 degrés assemblées pour donner une impression de continuité [8]. Concrètement, le service permet à tout un chacun de visualiser, sur Internet, toute partie de la voie publique et de naviguer virtuellement dans les rues de villes et de villages. Le projet exploite la technologie de la société "Immersive Media", qui permet de fournir une vue de la rue à 360 degrés en n'importe quel point donné de cette rue.

Ces photos affichent souvent des données à caractère personnel, comme des promeneurs dans une rue commerçante, des plaques d'immatriculation de voitures, des maisons, etc. C'est la raison pour laquelle cette application est soumise à la Loi vie privée.

En ce qui concerne la protection de la vie privée, Google a mis en oeuvre un certain nombre de mesures de nature à garantir le respect des droits des personnes dont les données apparaissent sur les images : photographies anciennes datant de quelques mois à quelques années et non représentatives de la voie publique actuelle, pas de visualisation en temps réel, un effet de flou est appliqué aux visages et aux plaques minéralogiques et éventuellement autre, sur demande, afin de les rendre méconnaissables. Un aperçu des mesures qui précèdent est disponible sur la page Google reprenant la politique de respect de la vie privée de Street View [9].

Cependant, entre 2008 et 2010, Google ne s'est pas contenté de photographier les routes, rues et maisons de nombreuses villes à travers le monde, il a également enregistré une panoplie de données telles que les coordonnées des réseaux sans fil, les adresses MAC des boîtiers associés, mais également des données de navigation plus privées, le tout à l'insu des citoyens. Tous les

réseaux sans fil se trouvant à portée ont été scannés et listés par ces voitures, bardées de capteurs, ce qui permet à présent à Google d'obtenir une géolocalisation avec une précision de 30 mètres.

Après s'être défendu dans un premier temps de tels agissements [10], Google a finalement admis la faute et fait amende honorable, ce qui n'a pas empêché l'entreprise d'être condamnée dans un grand nombre de pays. En Allemagne notamment, où l'autorité de protection des données de Hambourg a condamné, en avril 2013, Google à s'acquitter de 145 000 euros d'amende pour avoir collecté des données personnelles à l'aide de ses voitures, entre 2008 et 2010. Ce pays s'est toujours montré particulièrement frileux et méfiant quant aux procédés de Google. Le lancement de Street View s'était heurté à l'attachement des citoyens au respect de leur vie privée et le programme n'est d'ailleurs plus actualisé outre-Rhin, où une vingtaine de villes seulement a reçu la visite des Google cars, il y a plus de deux ans [11].

1.1.4 La SNCB Europe divulgue des informations personnelles sur son site

Le 22 décembre 2012, un internaute a découvert les informations personnelles de plus d'un million de personnes sur le site Internet de la SNCB Europe suite à une requête sur un moteur de recherche, à savoir Google. Cette liste contenait les nom, prénom, date de naissance, civilité, sexe, langue, adresse de messagerie électronique, adresse postale, numéro de téléphone mais aussi identifiants de connexion de clients belges, français ou britanniques. Ces informations auraient été accessibles sur le site Internet de la SNCB plusieurs semaines durant.

Les investigations réalisées au sein de SNCB Europe ont démontré qu' "une erreur humaine involontaire est à l'origine de la mise en ligne malencontreuse de ce fichier" [12]. Aucun chemin d'accès ne mène à ce document sur le site mais l'adresse URL de ce document, sans mot de passe pour la protéger, aurait été d'une certaine manière indexée par les moteurs de recherche.

Cette fuite d'informations met en avant une gestion irresponsable des données personnelles des usagers de la SNCB Europe en ne prenant aucune mesure pour garantir que celles-ci soient inaccessibles au public. Jusqu'à présent, la Commission vie privée a reçu plus de 1700 plaintes [13].

Quelques jours plus tard, des données militaires étaient également accessibles sur la toile. L'annuaire téléphonique de la Direction générale des ressources humaines de l'armée, un document de 37 pages en principe uniquement accessible sur le réseau intranet de la Défense, est consultable sur Internet et ce, grâce à la même méthode que pour la SNCB Europe, à savoir une recherche ciblée sur un moteur de recherche [14].

Ces deux faits divers illustrent parfaitement une mauvaise politique de protection des données de la part d'un gestionnaire de service mais aussi des utilisations détournées des moteurs de recherche. En effet, ceux-ci indexent la totalité d'Internet et "peuvent être utilisés par des personnes mal intentionnées pour mettre au jour les failles de sécurité de certains sites et les exploiter." [15].

1.1.5 Les cookies utilisés par le gouvernement américain

Tout internaute est-il un espion potentiel ? C'est du moins ce que semble penser la CIA (*Central Intelligence Agency*) qui espionne les visiteurs naviguant sur leurs sites Internet à l'aide de cookies persistants laissés sur leurs ordinateurs. En 2002, Daniel Leslie Brandt [4], activiste améri-

4. Daniel Leslie Brandt a mis en ligne les sites Google Watch et Wikipedia Watch, contenant des critiques sur le moteur de recherche et l'encyclopédie en ligne. Il est également le développeur de Scroogle, un proxy pour le moteur de recherche Google dont le but est de "nettoyer" les résultats de leurs publicités et d'empêcher le traçage de l'activité de l'utilisateur via des cookies [16]. Scroogle finit par fermer son service en février 2012 suite à des

cain très impliqué dans les problèmes de respect de la vie privée sur le Web, découvre la présence de ce procédé. Une fois informée de cette violation, la CIA déclara que ces cookies n'étaient pas créés intentionnellement et cessa leur mise en place [17].

En 2000, le gouvernement des États-Unis a mis en place des règles strictes concernant l'usage des cookies, et plus précisément les cookies persistants, après qu'il fut révélé que le bureau des politiques antidrogues de la Maison Blanche utilisait cette technologie pour suivre les internautes consultant en ligne les publicités antidrogues.

1.2 La vie privée

Le concept de vie privée est souvent perçu comme une notion vague, difficile à définir. Avant d'aborder le sujet de protection de la vie privée, de droit à l'anonymat ou de profilage de nos informations personnelles, il est nécessaire d'éclaircir cette zone d'ombre.

La vie privée relève de nos informations personnelles, cela va de soi. Ce terme est souvent utilisé pour poser une barrière, dénoncer un abus et finalement faire valoir son droit à la vie privée, la non divulgation des informations que nous jugeons sensibles. Mais où se situe la frontière entre le public et le privé ? Chacun a sa propre opinion, tout le monde n'a pas la même pudeur. Cette limite est d'autant plus difficile à définir que, dans le contexte du numérique, chaque information est conservée. Dans la vie quotidienne, les informations échangées oralement ne sont plus accessibles au moment suivant, il faut faire l'effort de noter ou d'enregistrer les conversations pour en retrouver le contenu. A l'exact opposé, la plupart des informations échangées sur Internet le sont par écrit. Il n'est nul besoin d'effort pour les conserver et il revient alors à l'internaute d'effectuer la démarche de les supprimer, ou de demander au responsable du site Internet de les supprimer. Et cette démarche est non seulement un effort à fournir, mais elle n'est en plus pas un automatisme.

Aujourd'hui, Internet est un outil utilisé quotidiennement par la plupart des gens. Que ce soit la communication par e-mail ou par messagerie instantanée, la discussion sur un forum, la recherche d'informations sur Internet ou l'achat en ligne, les ordinateurs, tablettes et autres *smartphones* sont entrés dans notre vie de tous les jours. Si bien que nous ne nous rendons plus compte que chaque information transmise est potentiellement enregistrée, répertoriée, traitée sans que nous y prenions garde.

Si l'internaute ne se méfie pas, ses données circulent d'une base de données à l'autre et se retrouvent dans un nombre toujours plus important de fichiers au risque de se retrouver rapidement harcelé de publicités. Des fenêtres publicitaires aux e-mails, cela peut devenir très envahissant. Plus grave, si un nombre suffisamment important d'informations à propos d'une personne sont en libre accès sur le net, quelqu'un de mal intentionné peut usurper son identité ou lui faire endosser la responsabilité d'actes ou de propos. Entre les réseaux sociaux, les blogs, les échanges d'e-mails, les recherches menées, les commentaires sur des forums de discussion, il n'est pas si ardu de rassembler un grand nombre d'informations relatives à une personne en ne connaissant au départ que son adresse de messagerie, son nom voir même un pseudonyme employé de manière régulière.

Il est également à remarquer que le développement croissant et l'omniprésence de la surveillance automatique et distante exercée sur l'individu menace sa capacité de développement personnel. En effet, la conscience de cette surveillance et des décisions semi-automatiques ou automatiques qui découleront, grâce notamment aux techniques de profilage et de recoupement

blocages causés par Google et des attaques de type DDOSS ou "déni de service".

d'informations, de la moindre déviance par rapport au comportement attendu lui impose une discipline implacable. Il en résulte une autocensure, éventuellement inconsciente, des individus, de crainte d'adopter un comportement qui serait perçu comme étrange par autrui, si les informations relatives à ce comportement venaient à être rendues publiques suite à la mise en oeuvre de certaines technologies de surveillance.

L'autocensure des individus pourrait aller jusqu'à les empêcher de participer à certaines activités collectives de la société civile, de peur que leur participation puisse être révélée et finalement leur nuire. C'est pourquoi la Cour constitutionnelle fédérale allemande estimait déjà en 1983 que si aucune mesure de protection de la vie privée n'est instaurée dans le cadre légal, le développement technologique auquel nous assistons depuis des années risque "de détruire non seulement nos chances de nous développer mais aussi le bien-être commun, car l'autodétermination est la condition fonctionnelle élémentaire d'une communauté démocratique libre fondée sur la capacité des citoyens d'agir et de coopérer" [5].

La Cour allemande considérait que le droit à la protection des données à caractère personnel est fondé sur les principes de la dignité humaine et du respect dû à l'autodétermination individuelle, définie comme étant le pouvoir de l'individu de décider lui-même de quand et dans quelle mesure une information relevant de sa vie privée peut être communiquée à autrui [18].

1.3 L'économie de l'attention

Ce concept, adopté en 1997 par Michael Goldhaber, désigne une nouvelle forme d'économie basée sur l'attention. Elle met en avant sa rareté et lui applique les modèles économiques, centrés sur l'attention comme ressource principale. Il y prévoit que les transactions financières actuelles seront remplacées par des transactions de l'attention, dans un monde où l'information est de plus en plus présente, et en quantité croissante, alors que l'attention des utilisateurs reste limitée. La tendance est donc inversée, la rareté ne se situe plus au niveau de l'information, mais au niveau de la capacité limitée de traiter cette quantité d'information. L'accès à l'information devant de plus en plus aisé, notamment de par la diminution de son coût, s'explique par la diffusion d'Internet. Nous parlons dès lors de société de l'information [6], et même de société de la connaissance [7] où les informations sont traitées, organisées, interprétées. Trois tendances font évoluer la problématique de l'attention [19] :

1. La dispersion de l'attention (Datchary - 2005) : un nombre sans cesse croissant de terminaux communiquent entre eux, parfois de manière simultanée, et de plus en plus souvent dans des situations de mobilité.
2. Le basculement du modèle économique sur Internet vers la publicité (Beuscart et Mellet - 2008) : de plus en plus présente, elle se retrouve sur la grande majorité des sites Internet.
3. Le risque de surcharge informationnelle (Eppler et Mengis - 2004) : risque auquel s'exposent de plus en plus les individus qui ont accès à un nombre exponentiel d'informations sur des supports variés tels qu'Internet, les intranets, les e-mails, etc.

5. Déclaration de la Cour constitutionnelle allemande, en 1993, lors de sa décision de l'inconstitutionnalité de certaines dispositions de la version révisée de la Loi de Recensement, suite aux recours introduits par diverses associations.

6. La société de l'information désigne un état de la société dans lequel les technologies de l'information jouent un rôle fondamental.

7. La société de la connaissance est un type de société où règne une forte diffusion des informations et qui agrège des savoirs pour faciliter la transmission des connaissances à ses membres.

L'information présentée résulte d'un compromis entre deux objectifs d'exploitation distincts. D'une part, comme les utilisateurs sont dans l'incapacité de tout voir et tout analyser, l'information est présentée de manière à économiser leur attention en leur montrant uniquement ce qu'ils veulent voir. D'autre part, obtenir l'attention d'un maximum de personnes revient à créer de la valeur, c'est le principe de la publicité. L'information est mélangée avec des annonces qui visent un objectif de marketing.

De cette théorie de l'économie de l'attention découle tout naturellement le principe de cibler au maximum les informations fournies pour conserver l'attention de l'utilisateur. Qu'il s'agisse de simplifier la tâche en effectuant un premier tri judicieux dans des résultats de recherche ou encore d'afficher une publicité pertinente qui attirera plus facilement l'attention. Or, le seul moyen de cibler les informations à fournir est de connaître le visiteur. Cela peut aller d'une connaissance très basique telle que sa région de résidence, son sexe, sa tranche d'âge à une connaissance approfondie incluant les centres d'intérêt passés et présents, les achats effectués sur Internet, les autres internautes avec qui la personne est en contact, etc. Le profilage saisit alors l'individu au plus proche de sa sensibilité, de par sa pertinence contextuelle en temps réel en fonction de l'endroit où il est, de ce qu'il fait ou lit.

1.4 Le Web social et l'exposition volontaire

A la collecte constante d'informations s'ajoute la tendance à la divulgation volontaire d'informations de la part de l'internaute. Depuis quelques années déjà, fleurissent les blogs personnels, les albums photo en ligne, les réseaux sociaux, etc. Tous ces outils forment une nouvelle dimension d'Internet, appelée le Web social, où la toile est considérée non plus comme un moyen de communication d'informations et de distribution de documents, mais bien comme un espace de socialisation où les utilisateurs peuvent se rencontrer et interagir.

Les sites Internet formant le Web social ont une finalité commune, à savoir se révéler aux autres, du moins en apparence. Il s'agit là d'un mécanisme visant à chercher l'approbation des autres. L'être humain construit sa vie sociale grâce au jugement que ses congénères lui portent. Etaler sa vie privée sur Internet, du moins en partie, fait partie de ce processus. Encore une fois, ces échanges d'informations peuvent paraître anodins ; mais lorsqu'il s'agit d'informatique, tout peut être enregistré et donc retrouvé. Une fois une information publiée sur Internet, son devenir est souvent incontrôlable. Supprimer la source que l'internaute a lui-même publiée ne peut lui garantir que personne n'en a fait une copie, ou même que l'information est entièrement supprimée. Peut-être y a-t-il eu une sauvegarde de sécurité entre-temps, peut-être les informations ont-elles été envoyées automatiquement vers d'autres destinations. Pensons à Facebook, où tout utilisateur peut configurer son compte pour être averti de chaque nouvelle publication de ses contacts. Dans ce cas, aucun retour en arrière n'est possible. Même si le message est supprimé ou modifié sur Facebook, les e-mails d'information ont déjà été envoyés.

1.5 Les paradoxes de la vie privée

Avec l'essor du Web social, beaucoup de personnes dévoilent leurs données à caractère personnel et même leurs pensées intimes. Susan B. Barnes [8] introduit la notion de *"privacy paradox"* [20] soit le **paradoxe de la vie privée** pour désigner la contradiction existant aux Etats-Unis entre d'un côté les adolescents qui divulguent volontiers leurs informations personnelles et sentiments et de l'autre côté les organismes gouvernementaux et commerciaux qui collectent ces données.

8. Professeur au département communication à l'institut des technologies de Rochester.

Les personnes communiquant sur les réseaux sociaux, particulièrement les adolescents, y trouvent une manière d'échanger des idées, leurs convictions et leurs goûts, notamment via les groupes Facebook, de se construire une identité, de rencontrer d'autres personnes. Seulement, la plupart de ces informations sont personnelles et peuvent être utilisées dans un cadre totalement différent de celui pour lequel les informations ont été publiées, chose à laquelle beaucoup ne pensent pas. L'on a vite fait d'oublier qu'Internet est public et que ces informations sont consultables par un grand nombre de personnes, voire par tout le monde. Selon James Ralph Beniger [9], "Plus les sociétés atteignent un haut degré d'organisation, plus les mécanismes de contrôle social se répandront, inévitablement". Susan B. Barnes précise que "pour obtenir des mécanismes de contrôle efficaces, une quantité considérable d'informations doivent être collectées". Voilà pourquoi les échanges d'informations sont aujourd'hui contrôlés par des organisations et des programmes. L'analyse des données échangées permet aux entreprises de glaner des informations à des fins publicitaires et de renforcer leur images de marque en adaptant leur politique commerciale aux profils visés.

En dehors de l'exploitation commerciale, il existe d'autres menaces plus sérieuses comme l'usurpation d'identité, le vol de données bancaires, l'exposition aux prédateurs sexuels, etc. Les informations personnelles les plus souvent exposées sur le Web social incluent les nom et prénom de l'internaute, sa date de naissance, mais également ses photos et de nombreux moyens d'entrer en communication avec celui-ci tels que l'adresse postale, l'adresse de courrier électronique ou un lien vers un journal personnel. Il est dès lors aisé, pour des personnes mal intentionnées, de repérer des "proies" et les contacter, en adoptant une éventuelle identité fictive. L'Internet rendant les communications moins directes, le sentiment de méfiance à l'égard des inconnus a tendance à laisser place à la curiosité et l'envie de faire de nouvelles connaissances.

Comme cela a été souligné ci-dessus, les problèmes relatifs au partage de données à caractère personnel touchent plutôt la tranche jeune de la population. Il semblerait que les "natifs du numérique" n'aient aucune pudeur à l'égard de leur vie privée. D'après le professeur Ravi Sandhu [10], ce phénomène serait comparable à l'attitude désinhibée avec laquelle les jeunes des années soixante abordaient la sexualité. De nombreux professionnels émettent l'hypothèse de la fin de la vie privée sur Internet. Citons Susan B. Barnes ("A privacy paradox : Social networking in the United States"), Simson Garfinkel ("Database Nation : the death of privacy in the 21th century") ou encore Jean-Marc Manach ("La vie privée, un problème de vieux cons ?").

Il existe un **autre paradoxe**, relatif aux résultats des recherches sur Internet. Comme spécifié dans la section traitant de l'économie de l'attention, les publicités sont ciblées et les informations fournies par les moteurs de recherche sont filtrées. L'internaute voit ses recherches triées selon certains critères, sans en connaître les détails et surtout, sans en avoir le choix. Il ne voit que ce que l'on veut bien lui montrer. Cet aspect discriminant non visible et non intelligible du profilage s'opère en fonction de son pays, de sa langue, et qui sait encore, de son métier ou ses centres d'intérêt. Cependant, ce filtrage des informations pertinentes peut également faciliter la recherche d'information et réduire les efforts intellectuels de l'utilisateur.

A force de lecture, nous pourrions relever un **dernier paradoxe**. Nombre de personnes se soucient de leur vie privée et du fait qu'elles restent, justement, dans un cadre privé. Elles sont conscientes des risques et désagréments auxquels elles s'exposent mais pourtant elles sont peu nombreuses à se protéger réellement. Attendent-elles que l'Etat mette en place une protection efficace ? Peut-être ces personnes sont-elles tout simplement incapables de se protéger car elles

9. Professeur de communication et sociologie à la *USC Annenberg School for Communication Journalism* , auteur du livre *The Control Revolution : Technological and Economic Origins of the Information Society.*

10. Professeur à l'université du Texas à San Antonio, responsable de l'Institut de la cyber sécurité.

ne connaissent que peu les techniques utilisées pour collecter leurs données et les moyens d'y échapper. Nous répondrons à ces deux questions en détaillant, dans ce document, les techniques les plus répandues sur le net.

1.6 La loi "vie privée"

D'après le "Baromètre de la société d'information 2013" publié par le SPF Economie, 78% des ménages Belges disposent d'un ordinateur connecté à Internet. Selon Facebook, en 2013, 5.013.860 Belges seraient inscrits sur le réseau social, soit 48% de la population. Les Belges sont donc d'importants utilisateurs d'Internet et plus particulièrement des réseaux sociaux. Ils ne sont pas non plus en reste concernant les achats en ligne. Pour régir tout ce transit d'informations, une loi existe, la loi du 8 décembre 1992, appelée également "Loi relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel" ou plus simplement "Loi vie privée". Celle-ci identifie comme données à caractère personnel "toute information concernant une personne physique identifiée ou identifiable". Elle précise cette définition par le fait qu' "est réputée identifiable une personne qui peut être identifiée, directement ou indirectement, notamment par référence à un numéro d'identification ou à un ou plusieurs éléments spécifiques, propres à son identité physique, physiologique, psychique, économique, culturelle ou sociale.". Toute donnée personnelle ou professionnelle est donc sujette à cette loi, que ce soit une photo, un numéro de téléphone, une adresse e-mail, un code, etc. Cette loi a donné naissance à un organe de contrôle indépendant appelé la Commission de la protection de la vie privée (CPVP) dont le rôle est de "veiller à ce que les données à caractère personnel soient utilisées dans le respect de la loi vie privée, avec le soin et les précautions qui s'imposent, de manière à préserver la vie privée des citoyens." [11]

Au niveau européen, il existe une directive [12] décrivant le régime général de la protection des données à caractère personnel et dont le but est d'harmoniser les règles de protection des données personnelles sur tout le territoire de l'Union européenne. Cette directive comporte néanmoins deux inconvénients majeurs à son application. Tout d'abord, de par sa nature, cette directive ne peut être appliquée directement au niveau national. Chaque Etat membre a donc dû la transposer au sein de sa législation, se réservant au passage quelques libertés d'interprétation. Ceci conduit à une législation finalement sensiblement variable d'un pays à l'autre, compliquant l'arbitrage en cas de conflit avec une société internationale possédant des sièges dans plusieurs pays européens. Le deuxième inconvénient de cette directive est dû à sa vétusté. La gestion de la vie privée a beaucoup évolué depuis 1995, notamment suite à l'essor d'Internet et la naissance des réseaux sociaux et plus généralement du Web social. C'est pourquoi la Commission européenne a jugé nécessaire d'actualiser la législation en rédigeant, le 25 janvier 2012, une proposition de règlement qui, une fois adoptée, devrait remplacer la directive existante. Cette proposition vise donc à adapter, harmoniser et renforcer le cadre légal existant, avec l'avantage d'être directement applicable dans les 27 Etats membres ainsi qu'en Norvège, en Islande et au Liechtenstein. De par sa qualité de loi européenne, elle primera sur toute législation nationale, y compris la loi vie privée de Belgique.

11. La loi vie privée, consultable sur le site de la Commission vie privée (www.privacycommission.be).
12. Directive 95/46/CE du Parlement européen et du Conseil du 24 octobre 1995.

1.7 Les droits des utilisateurs

Afin de comprendre quelles données seront collectées, pourquoi et par qui, et de pouvoir gérer les informations renseignées, la loi vie privée prévoit que les utilisateurs aient les droits, notamment, de connaître [13] :

- le nom et l'adresse du responsable du traitement et, le cas échéant, de son représentant
- les finalités du traitement
- l'existence d'un droit de s'opposer, sur demande et gratuitement, au traitement de données à caractère personnel les concernant et envisagés à des fins de marketing direct
- les destinataires ou les catégories de destinataires des données
- le caractère obligatoire ou non de la réponse ainsi que les conséquences éventuelles d'un défaut de réponse
- l'existence d'un droit d'accès et de rectification des données les concernant ; sauf dans la mesure où, compte tenu des circonstances particulières dans lesquelles les données sont obtenues, ces informations supplémentaires ne sont pas nécessaires pour assurer à l'égard de la personne concernée un traitement loyal des données

1.8 Les devoirs des gestionnaires de systèmes d'information

De son côté, le gestionnaire du système collectant des données doit s'assurer de ne recueillir que les informations strictement nécessaires aux finalités exposées à l'utilisateur et ce pour une période maximum correspondant à la durée requise par ces finalités. Il est de son devoir, une fois ce délai dépassé, de supprimer ou brouiller les informations des utilisateurs afin de les rendre non identifiables. Pour cela, les données à caractère personnel doivent être :

- traitées loyalement et licitement
- collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement de manière incompatible avec ces finalités
- adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont obtenues et pour lesquelles elles sont traitées ultérieurement
- exactes et, si nécessaire, mises à jour
- conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire à la réalisation des finalités pour lesquelles elles sont obtenues ou pour lesquelles elles sont traitées ultérieurement.

Toutes les données personnelles n'ayant pas le même degré d'intimité, certaines d'entre elles, dites données sensibles, sont régies par une règle spécifique. Plus précisément, les données dont le traitement est interdit sont :

- les données à caractère personnel qui révèlent l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques, l'appartenance syndicale, ainsi que le traitement des données relatives à la vie sexuelle
- les données à caractère personnel relatives à la santé
- les données à caractère personnel à caractère juridique (relatives à des litiges soumis aux cours et tribunaux ainsi qu'aux juridictions administratives, à des suspicions, des poursuites ou des condamnations ayant trait à des infractions, ou à des sanctions administratives ou des mesures de sureté).

13. Exemple de charte de confidentialité : http ://global.joinfite.org/fr/privacy-and-legal

Le gestionnaire du système doit également assurer la protection des données qui lui ont été confiées. Bien qu'aucune protection ne soit infaillible, il se doit de prendre toutes les précautions utiles, en fonction de la nature des données et des risques présentés par le traitement, pour préserver la sécurité des données. Il devra respecter les règles suivantes :

- protéger son système contre tout accès illégitime, extérieur comme intérieur.
- assurer l'inaltérabilité des données enregistrées.
- enregistrer les mots de passe des clients sous leur forme chiffrée et non le mot de passe original.
- respecter les règles relatives à l'enregistrement des données bancaires. Celles-ci prévoient notamment l'interdiction de stocker en clair les numéros de cartes des clients.

Il s'agit donc d'une obligation de moyens, non de résultats. Ce n'est pas le cas aux Etats-Unis [14] ou en Allemagne par exemple, où tout manquement à la sécurité des informations personnelles stockées par un système d'information se doit d'être signalé à un organisme légal de contrôle responsable, ainsi qu'aux personnes directement concernées par la fuite, perte ou altération de données.

1.9 Les devoirs des fournisseurs d'accès à Internet

L'accès à Internet n'est pas gratuit, et pour cause toute une infrastructure doit être mise en place, de la pose des câbles et leur entretien à la gestion des évolutions technologiques, en passant par la gestion des communications transitant sur le réseau. Ces dernières nécessitent, entre autres, la distribution d'adresses IP aux clients souhaitant se connecter à Internet, dont les détails seront abordés dans le prochain chapitre. Une fois la communication établie, les requêtes sont relayées de serveur en serveur pour atteindre le destinataire. Tout ce travail est assuré par des organismes, généralement des entreprises, appelées fournisseurs d'accès à Internet, souvent abrégé en FAI.

Les FAI sont les seules organisations légalement autorisées et tenues de connaître l'identité de la personne possédant une adresse IP, ils doivent donc enregistrer les différentes allocations d'adresses IP attribuées à leurs clients. Ces enregistrements permettent, en cas d'enquête judiciaire, de remonter des informations de connexion, liées à une activité durant une session de navigation, vers la personne physique qui les a initiées [21].

Pour chaque communication passée, les fournisseurs doivent donc être capables d'identifier l'utilisateur mais pas seulement. Sont soumis au même suivi le terminal utilisé, ainsi que le type et la durée des communications. Ces données sont à conserver pour une durée de un an exactement [22], selon la directive européenne 2000/31/CE du 8 juin 2000 relative au commerce électronique. Cette directive est implémentée, en France, par la LCEN (Loi pour la Confiance dans l'Economie) [23].

Toute entreprise qui fournit un accès à Internet à ses employés ou à ses visiteurs est considérée comme un fournisseur d'accès à part entière et est soumise aux mêmes obligations légales de journalisation des connexions que celles assignées aux FAI. Bien que la loi ne le stipule pas clairement, un jugement établi contre l'établissement bancaire BNP-Paribas en 2005 fait jurisprudence [24].

14. Selon les *Security breach notification laws*.

1.10 Le profilage

La législation existe donc et un organisme de contrôle est présent pour assurer son application, mais celle-ci ne couvre pas le profilage anonyme qui consiste à collecter des informations relatives à l'internaute mais qui ne sont pas des données à caractère personnel tel qu'entendu par la loi et qui ne permettent donc pas d'identifier la personne concernée. Citons par exemple la tranche d'âge, la région, les centres d'intérêt supposés, le sexe, etc. Autant d'informations qui permettent de cibler les offres et publicités affichées mais qui ne permettent pas, directement ou indirectement, de retrouver l'identité officielle de la personne concernée. Ceci est d'autant plus vrai que les utilisateurs sont, la plupart du temps, placés au sein d'une catégorie d'utilisateurs aux caractéristiques similaires selon certains critères dépendant bien entendu de l'objectif de l'organisme qui exploite les données, par exemple le sexe, des catégories d'âge, la nationalité, etc. Ces catégories d'utilisateurs sont appelées profils et la technique qui consiste à situer chaque personne au sein des différentes catégories s'appelle le profilage.

Cette technique, basée sur l'étude statistique des caractéristiques d'un ensemble d'utilisateurs, permet aux prestataires de services de personnaliser leur offre. Cela peut aller de l'esthétique au fonctionnel ou encore à l'orientation des publicités. Il est beaucoup plus aisé de configurer une offre en prévoyant un nombre limité de profils que de tenter de personnaliser un service en fonction de chaque visiteur.

Le profilage est une technique qui a le potentiel de profiter aux utilisateurs mais également de nuire à leur confort d'utilisation d'Internet. Il existe, en effet, des effets pervers à l'utilisation de profils, tout dépend des objectifs de l'organisme concerné et de la manière dont il exploite ses informations. Citons par exemple la limitation des fonctionnalités en fonction du profil de l'internaute. Dans ce cas, il ne sera peut-être même pas au courant de cette limitation et perdra l'accès à certaines informations. Ces limitations peuvent être volontaires de la part de l'organisme, mais elles peuvent aussi être involontaires, plutôt une conséquence de l'orientation de l'information. On tente de présenter les informations qui ont le plus grand potentiel d'intéresser l'utilisateur et au final on en oublie qu'une autre information aurait pu lui être utile. Ce choix peut être motivé par les limitations humaines concernant la quantité d'informations traitables, évoqués dans la théorie de l'économie de l'attention.

Le profilage est très présent sur Internet, et malheureusement se déroule souvent à l'insu de l'utilisateur. Pour éviter les éventuels désagréments qui en découlent, il existe de nombreuses techniques permettant de s'en protéger, ou du moins d'en limiter l'impact.

Chapitre 2

Les données et le profilage

Nous savons à présent que les données personnelles ont beaucoup de valeur. Elles sont le moteur du profilage qui, après traitement, permet de toucher l'internaute dans le contexte précis de ce qu'il est et de ce qu'il fait. Nous savons également combien il est ardu de conserver ces informations confidentielles sur un Internet où le Web social est de plus en plus présent. Il incombe donc à chacun d'appliquer une politique préventive en mettant en place des techniques défensives à la source, c'est-à-dire au niveau des ordinateurs personnels. Le meilleur moyen de limiter la répartition des informations est évidemment de réguler leur divulgation en choisissant consciemment, au cas par cas, quelle donnée partager, avec qui et en quelles circonstances.

Pour ce faire, il est primordial de connaître un minimum les techniques utilisées pour collecter les informations des internautes, que ce soit de manière explicite ou, au contraire, totalement transparente et donc pour beaucoup inconnue. Sur la toile, chaque action ou inaction a des conséquences.

Il serait impossible de parcourir la totalité des moyens de collecte des données, nous nous limiterons donc aux plus répandus pour un utilisateur lambda en considérant pour chacun la nature de l'information collectée.

Certains points sont incontournables car ils concernent tous les internautes, quelle que soit leur activité sur Internet, il s'agit de : l'adresse IP, les cookies, les en-têtes HTTP et l'historique de navigation.
D'autres touchent une majorité de personnes en recouvrant les activités les plus répandues sur Internet, tels que : les moteurs de recherche, les communications électroniques, les services de localisation et les formulaires en ligne.
Une troisième catégorie aborde les secteurs en plein essor : le Web social, le commerce électronique et le *Cloud computing*.
Finalement, les informations sur les mobiles et les logiciels espions sont des points plus spécifiques mais concernent des atteintes directes à la vie privée et trouvent de ce fait leur place dans ce tour d'horizon.

Lorsqu'une catégorie est trop vaste, son représentant le plus courant sera présenté. Par exemple, le *Cloud computing* recouvrera les offres de stockage en ligne via un disque virtuel mais également tout service proposant la sauvegarde en ligne des informations et paramètres de ses clients ; les commerces électroniques regroupent quant à eux tout site Internet dont l'activité de l'utilisateur peut être exploitée en tant que telle à des fins publicitaires, regroupant de ce fait les grandes plateformes comme les petits commerces spécialisés, les sites de jeux en ligne, les sites de réservations de vacances, etc.

2.1 L'adresse IP

L'adresse IP correspond à l'adresse de l'ordinateur, son identité. Toute machine désirant communiquer avec d'autres appareils sur un réseau local ou sur Internet doit obligatoirement posséder une adresse. Celle-ci permet aux communicants de savoir de qui provient le message reçu et donc à qui envoyer la réponse. L'adresse IP, provenant de l'anglais *Internet Protocol*, est attribuée par le fournisseur d'accès à Internet, auprès duquel l'internaute a souscrit un abonnement [25]. Les FAI sont les seules organisations légalement autorisées et tenues de connaître l'identité de la personne possédant une adresse IP. Les autres organisations sont, la plupart du temps, limitées à la connaissance de l'adresse IP et constituent donc un profil anonyme dans lequel elles regroupent toutes les informations qu'elles ont pu collecter, mais sans connaître l'identité de la personne physique exploitant cette adresse.

L'attribution d'une adresse a une durée limitée à 36 heures, pour les abonnements ayant un système d'IP dynamique, ce qui est le cas le plus courant. En général, seuls les possesseurs de serveurs souscrivent à un système d'IP fixe, car il est important que leur machine ne change pas d'adresse, pour que chaque ordinateur sache quelle adresse contacter pour entrer en communication avec eux. En théorie, l'ordinateur d'un internaute est donc méconnaissable entre deux sessions de navigation espacées de 36 heures ou plus. Il existe cependant des techniques permettant de remédier à cela, techniques que nous aborderons prochainement.

Il est important de savoir que les adresses IP ne sont pas distribuées au hasard, mais au terme d'une organisation hiérarchique [26]. L'organisme en haut de la pyramide est l'IANA (*Internet Assigned Numbers Authority*) qui est en charge de la gestion de l'espace d'adressage IP pour Internet. Elle remplaça le NIC (*Network Information Center*) en 1972 et fut en charge de la distribution des adresses jusqu'en 1990. Cette année, la RFC 1174 proposa un modèle de distribution hiérarchique des adresses, illustré à la figure 2.1, où l'IANA délèguerait la gestion de blocs d'adresses à des entités appelées RIR (*Regional Internet Registry*).

Ils sont actuellement au nombre de cinq, à savoir l'AfriNIC, l'APNIC, l'ARIN, la LACNIC et la RIPE NCC, répartis sur le globe. De nos jours, les RIR ne distribuent plus non plus d'adresses aux utilisateurs finaux mais délèguent à leur tour l'attribution des adresses à d'autres entités appelées LIR (*Local Internet Registry*), qui sont habituellement des FAI. Certains pays possèdent un unique NIR (*National Internet Registry*) qui est en charge de l'attribution des blocs d'adresse IP aux LIR pour un pays ou une zone économique donné.

Il y a donc encore un niveau intermédiaire supplémentaire entre l'IANA et les utilisateurs finaux. L'APNIC possède sept NIR gérant les allocations d'adresse pour l'Indonésie, la Chine, le Japon, la Corée, Singapour, Taiwan et le Vietnam. La LACNIC, quant à elle, en possède cinq, gérant l'Argentine, la Bolivie, le Chili, le Mexique et le Brésil. Une fois qu'il a reçu la charge d'adresses IP, le LIR décide lui-même quelles adresses sont utilisées pour ses propres équipements et quelles adresses sont distribuées à ses clients.

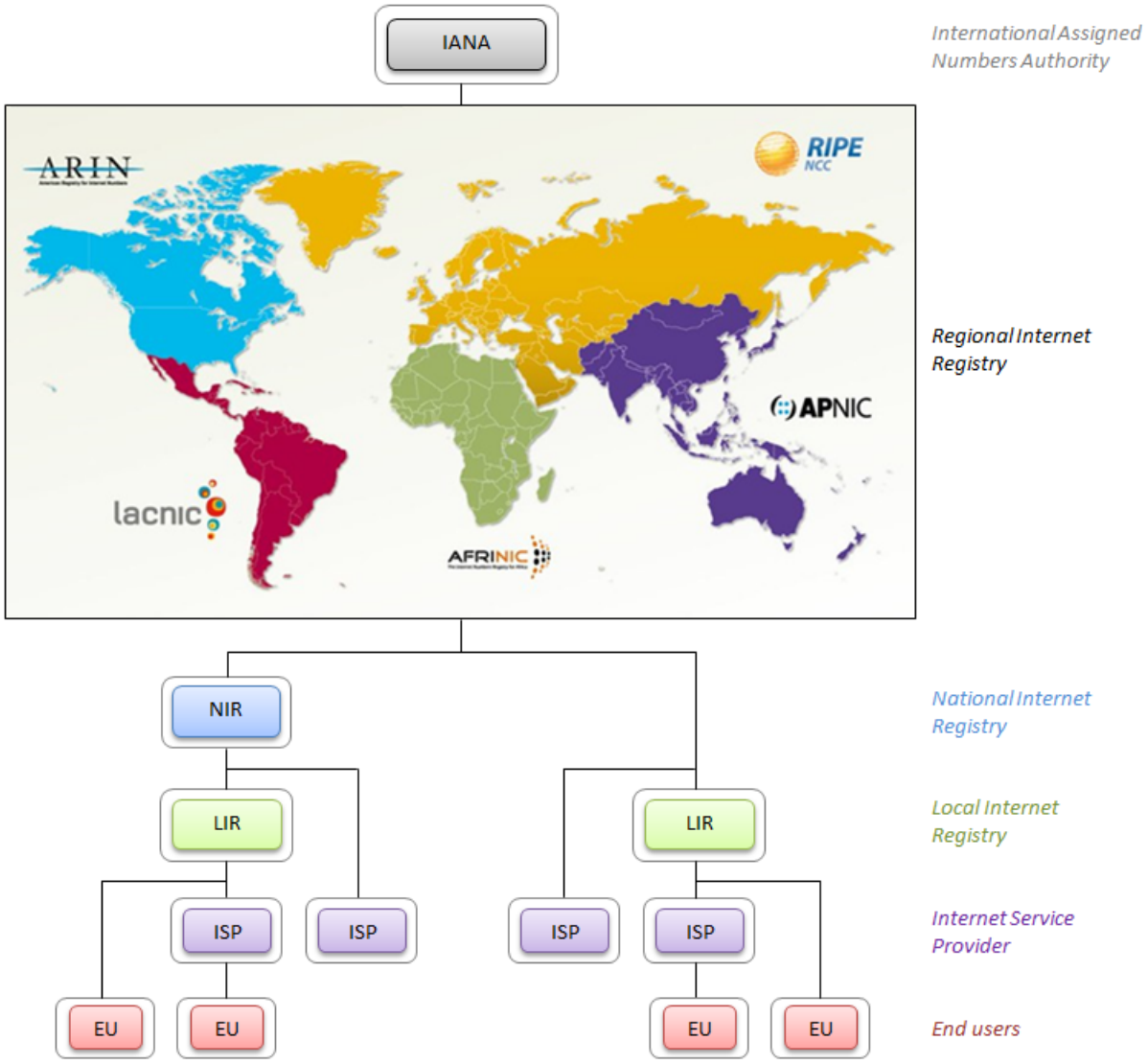


FIGURE 2.1 – Modèle hiérarchique représentant les divers organismes assurant la distribution des adresses IP à travers le monde

Comment cette information est-elle interceptée ?

L'adresse IP est présente dans l'en-tête de chaque paquet de données envoyé sur Internet. Il est donc impossible de la cacher. Sans elle, aucun point de communication n'est possible. Certains sites Internet tels que www.adresseip.com offrent d'indiquer à l'internaute sa propre adresse IP.

Quelle information est dévoilée ?

L'adresse IP révèle l'identité réelle de la personne ou l'organisation à qui elle est attribuée. En cas d'adressage dynamique, seul le FAI a accès à cette information. En cas d'adressage statique, avec une IP qui ne change jamais, tout le monde peut connaître l'identité de l'organisation grâce au service WHOIS [1]. En effet, lorsqu'une demande d'adresse fixe est adressée à un FAI, celle-ci nécessite également de fournir d'autres informations, qui seront accessibles librement en consultant les bases de données des RIR, notamment le nom ainsi que le numéro de téléphone et

1. Un WHOIS, littéralement "qui est ce", est un service d'accès à une base de données de contacts relative à une extension de noms de domaines.

l'adresse de courrier électronique auxquels le propriétaire est joignable. Le service WHOIS publie les mêmes informations pour une recherche basée sur un nom de domaine.

L'emplacement approximatif de l'appareil, par géolocalisation, s'effectue par exploitation des données d'enregistrement des adresses IP transmises aux organismes en charge de la gestion des adresses IP dans le monde, c'est-à-dire les cinq RIR [25]. Il est ainsi possible de déterminer le pays et le fournisseur d'accès Internet ayant reçu la charge de l'adresse IP. Les FAI contribuent également à enrichir la base de données avec des informations plus précises sur l'attribution des adresses qu'ils ont effectivement effectué. Grâce au service WHOIS des RIR, Il est donc possible de connaître la région, la ville, et donc la langue supposée de n'importe quel internaute. Cette information n'est toutefois pas totalement fiable, car elle dépendra du sérieux des FAI. Les adresses de type dynamique changeant de propriétaire toutes les 36 heures, les FAI doivent en référer régulièrement à leur RIR. De plus, certains FAI ne fournissent que la région pour laquelle l'adresse a été attribuée, et non la ville exacte. Notons comme exemple d'application que de certains sites proposent d'effectuer des requêtes aux services WHOIS et de situer le propriétaire actuel de l'adresse IP sur une carte.

Concernant les appareils mobiles, la localisation est bien plus compliquée, puisque l'internaute peut être en déplacement. Par exemple, en contactant le service WHOIS du RIR européen RIPE depuis un *smartphone*, le résultat est que l'on n'observe aucune information concernant l'utilisateur. Le service WHOIS ne possède que le nom du centre mobile de l'opérateur ainsi que l'adresse postale de son siège central. Sont également présents les noms et numéros des personnes de contact du centre. Les terminaux mobiles disposent toutefois de moyens de géolocalisation qui leur sont propres et ne se basent pas sur l'adresse IP. Ces techniques seront expliquées à la section 2.7 "Les services de localisation".

L'annexe A contient des exemples de géolocalisation par adresse IP, ainsi que de requêtes au service WHOIS via un nom de domaine et une adresse IP d'un terminal mobile.

2.2 Les cookies

L'adresse IP, bien qu'elle permette d'identifier un internaute de manière unique, est éphémère. Renouvelée toutes les 36 heures, comment s'y prennent les systèmes d'information pour reconnaître un visiteur ? Les cookies de navigation ou encore témoins de connexion, remplissent cette fonction. Il s'agit de petits fichiers texte stockés par le navigateur web sur le disque dur de l'ordinateur [27]. Ces fichiers sont créés, consultés et mis à jour par le navigateur de l'internaute, à la demande du serveur Web. Ce dernier envoie le cookie en tant qu'en-tête HTTP [2] et le navigateur le renvoie, inchangé, à chaque fois qu'il accède au dit serveur, introduisant un état dans la transaction HTTP. Sans l'utilisation des cookies, chaque accès à une page Web est un évènement isolé, indépendant des autres requêtes faites auprès du même serveur.

Le cookie contient généralement un identifiant unique appelé identifiant de session et sert à enregistrer des informations spécifiques au visiteur ou encore sur son parcours sur le site Web. Les informations enregistrées sont libres. Il peut donc s'agir des préférences de l'utilisateur par rapport au site, telles que la langue, un style d'affichage, des champs à pré-remplir à la prochaine visite, etc. La fonction pour laquelle les cookies furent inventés est d'ailleurs la gestion du contenu d'un panier d'achat. De nos jours, les cookies contiennent la plupart du temps uniquement l'identifiant de session, qui réfère à une base de données relative au serveur. Ceci permet d'alléger l'échange de données entre serveur et ordinateur client, puisqu'il suffit de propager cet

2. HTTP est un protocole de transfert Hypertexte gérant le transfert des pages Internet.

identifiant comme unique contenu du cookie. Cette technique supprime la contrainte liée à la taille limitée des informations qu'un cookie peut stocker et, finalement, permet au gestionnaire du serveur Web d'effectuer des statistiques sur l'ensemble de leurs visiteurs en consultant sa base de données. Chose qui était impossible lorsque les informations étaient éparses sur les ordinateurs clients. Un cookie peut également être utilisé pour une authentification par nom d'utilisateur et mot de passe ou pour maintenir une session, qui consiste à mémoriser l'état, la mémoire des évènements effectués par le visiteur du site durant sa navigation de page en page. Ces petits fichiers ont une durée de vie limitée, fixée par le concepteur du site. On parle de **cookies de session** si la date d'expiration n'est pas indiquée, le fichier est alors supprimé lors de la fermeture du navigateur. Spécifier cette date est donc un moyen de faire survivre le cookie à travers plusieurs sessions de navigation. Ce type de cookie est appelé **cookie permanent**. La loi concernant les cookies permanents stipule que ces derniers doivent être supprimés à la fin d'une période de 6 mois maximum. Notons encore, qu'une fois le petit fichier créé, le serveur n'a plus aucun moyen de connaître la date d'expiration. Il ne peut donc pas savoir si le cookie est en fin de vie ou non et si les infos vont disparaître ou pas [28].

Les cookies ne sont pas exécutables, ils ne peuvent donc pas écrire ou lire d'information provenant de l'ordinateur de l'utilisateur, contrairement aux virus et autres programmes d'espionnage. Leur fonction première est de simplifier la vie des visiteurs et leur présenter des informations plus pertinentes, comme la personnalisation du site visité. Malheureusement, ils ont été parfois détournés de cette fonction et servent à pister les activités d'un internaute lors d'une séance de navigation. Le pistage peut même avoir lieu à travers différents sites à l'aide de **cookies "tierces parties"** ou de pixels espions. Les images, publicités et autres objets contenus dans une page Web peuvent provenir de serveurs différents de celui hébergeant la page consultée. Tous ces objets sont téléchargés par le navigateur lors de l'affichage de la page. On appelle cookies "première partie" ceux qui sont mis en place par le domaine inscrit dans la barre d'adresse du navigateur, et cookies "tierce partie" ceux provenant d'un domaine différent, et généralement utilisés par les entreprises de publicités afin suivre un utilisateur sur plusieurs sites, collecter et recouper des informations sur ses habitudes pour cibler leur publicité. Déjà en 1996, soit deux ans après la première apparition des cookies sur Internet, le groupe de travail de l'IETF (*Internet Engineering Task Force*) en charge de spécification formelle des cookies détermina que les cookies tierce partie étaient une menace considérable à la protection de la vie privée.

Il existe une forme dérivée de cette technologie. Les **cookies LSO** (*Local Shared Object*) ou objets localement partagés [29] sont des objets créés de la même manière que les cookies, mais en exploitant la technologie Flash de la société Adobe. Ces objets sont donc gérés indépendamment des cookies, ce qui apporte quelques avantages pour les exploitants du site Internet. La capacité de stockage d'informations est plus importante, leur utilisation est commune aux éventuels multiples navigateurs installés et leur utilisation est possible alors que le navigateur du visiteur refuse les cookies. Ce dernier point n'est évidemment pas à l'avantage de l'utilisateur qui doit être conscient de l'existence de ce mécanisme lors de la configuration des paramètres de confidentialité de son navigateur. Et inversement, savoir que le blocage des objets locaux partagés n'influe pas sur la gestion des cookies.

L'annexe B illustre un exemple d'échange de cookies.

Comment cette information est-elle interceptée ?

Tout comme l'adresse IP, les cookies ne sont pas compliqués à récupérer. Plus encore, leur vocation est de pouvoir être consultés par le serveur Web qui en a initié la création. Lors de la visite du site Internet, le navigateur de l'internaute consulte les cookies dont il a la gestion. S'il

en possède un, créé par le site actuellement accédé et qui n'est pas expiré, le navigateur l'envoie au serveur en tant qu'en-tête HTML. Un serveur Web ne peut donc accéder qu'aux données qu'il a lui-même fournit initialement au navigateur dans un cookie.

Quelle information est dévoilée ?

Le système des cookies est avant tout une technique de reconnaissance de l'internaute qui se présente sur le site. Il n'y a donc pas de collection de données autre que celle relative aux activités des sessions. Un site peut donc connaître les dates et heures de chacune de vos visites, les pages visitées et leur ordre de consultation. Cependant, comme nous l'avons vu, des cookies peuvent être créés de manière plus subtile par des éléments présents sur les pages consultées. Une bannière publicitaire ou même l'affichage d'un unique pixel appelé d'ailleurs pixel espion, fait donc appel à un serveur tiers qui, lui aussi échange des cookies avec l'ordinateur. Si ce serveur est référencé par plusieurs sites consultés par l'internaute, il a donc connaissance de l'activité de l'internaute, non limitée à un site en particulier cette fois. En fonction des pages visitées par l'internaute, il est donc tout à fait possible de connaître ses centres d'intérêt et d'en déduire toutes informations possibles telles que son âge supposé, son sexe, son sport préféré, etc.

Notons que les cookies première partie ont la plupart du temps comme objectif l'amélioration de l'expérience utilisateur et éventuellement le calcul de statistiques relatives à l'utilisation du site Internet. Les cookies tierce partie en revanche ont comme principal objectif le ciblage des publicités affichées, tel que représenté à la figure 2.2.

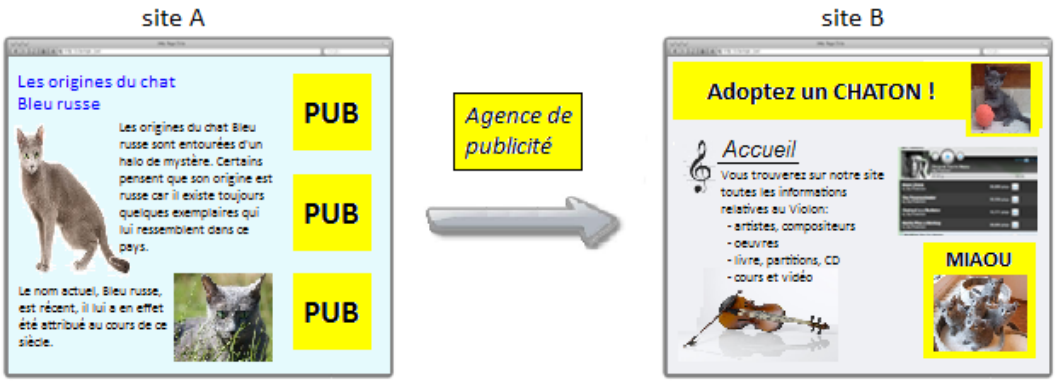


FIGURE 2.2 – Les agences de publicité utilisent les cookies tierce partie pour traquer le comportement d'un utilisateur sur un site A et afficher des publicités basées sur ce comportement sur un site B

La capacité des cookies tierce partie à collecter et exploiter les données de navigation viole la vie privée des utilisateurs pour plusieurs raisons :

- Impossible à détecter. L'utilisation de cookie tierce partie se passe en arrière-plan. L'utilisateur remarque généralement l'effet de ce pistage via la publicité ciblée, quand il est déjà trop tard.
- Se passe du consentement de l'utilisateur. Le serveur à l'origine de la création de tels cookies n'avertit pas le visiteur et lui demande encore moins son consentement. Les rares à s'en soucier utilisent généralement des tournures de phrase visant à induire l'utilisateur en erreur. Le site Web demandera par exemple "Voulez-vous que du contenu pertinent soit affiché en fonction de vos habitudes ?" au lieu de "Voulez-vous que les publicités soient ciblées en fonction de vos informations personnelles ?".
- Difficile à prévenir. Leur utilisation est hors du contrôle du site légitime visité par l'internaute, qui référence le serveur espion sans nécessairement être au courant de l'utilisation

de cette technique d'espionnage. Il faudrait donc éviter tout site comportant des bannières publicitaires, en considérant que les sites visités ne transmettent pas eux même vos informations de navigation à des tiers.
– Peut conduire à des situations indésirables. Le principe de la publicité ciblée consistant à collecter des informations à un moment donné pour la restituer indirectement à un autre moment. Certains contenus visités par l'internaute peuvent être particulièrement sensibles. Or, même en utilisant la navigation privée offerte par les navigateurs actuels qui assurent qu'aucun historique n'est conservé durant ce type de navigation, l'utilisateur ne peut se prémunir contre l'affichage inopiné de ses actions passées.

Il est un défaut de précision notable dans l'exploitation des cookies, à savoir qu'ils ne permettent pas de gérer la subtilité induite par l'utilisation de multiples navigateurs. En effet, chaque navigateur conserve les cookies dont il a la charge dans une unité de stockage qui lui est spécifique. Un serveur n'est donc pas capable d'identifier un utilisateur, mais plutôt une combinaison ordinateur, compte d'utilisateur et navigateur Web. En changeant l'un de ces trois paramètres, le serveur ne peut reconnaître un visiteur.

2.3 Les en-têtes HTTP

De nombreuses informations sont accessibles et fournies à tous les serveurs auxquels l'ordinateur de l'internaute envoie une requête [25]. Ces variables sont contenues dans l'en-tête de chaque requête partant de l'ordinateur. A l'origine, elles permettaient d'adapter le rendu du site Internet en fonction de la machine du client et donc d'enrichir son expérience sur le site. Citons l'adaptation de l'affichage à la configuration matérielle et logicielle du visiteur, l'affichage de ces pages dans sa langue, proposer directement les téléchargements adaptés au système d'exploitation et à sa version, etc. De nos jours, la plupart de ces informations sont devenues inutiles pour le confort de navigation car, les protocoles ayant évolué vers plus de standardisation, il n'est plus nécessaire, lors de la conception d'un site Internet, de se soucier des caractéristiques matérielles et logicielles des clients. Ces informations restent présentes pour des raisons historiques mais aussi statistiques. Elles ont par ailleurs été enrichies au fur et à mesure de l'avancée des versions des différents navigateurs.

Si certaines informations sont purement techniques et ne présentent actuellement qu'un intérêt statistique voire une absence totale d'intérêt, d'autres peuvent révéler de précieuses informations au sujet du visiteur, particulièrement lorsqu'elles sont recoupées avec des données d'espionnage obtenues par d'autres procédés. En somme, les informations présentées ici viennent compléter le profil public que d'aucuns dressent au sujet des internautes.
Parmi les informations fournies, citons :
◇ l'adresse IP
◇ le navigateur utilisé ainsi que sa version et sa langue
◇ le système d'exploitation utilisé ainsi que sa version
◇ la résolution d'écran et la qualité des couleurs
◇ le nom de l'ordinateur
◇ le port utilisé pour communiquer avec le serveur
◇ le nom de l'hôte (*hostname*) utilisé pour accéder à Internet, révélant le nom du FAI
◇ les pages accédées précédemment par le navigateur durant la session de navigation, donc d'où vient le visiteur
◇ la liste des plugins installés pour le navigateur ainsi que la liste des types de fichiers qu'il est capable de lire
◇ le fait que Java et/ou JavaScript soient activés ou non

Comment cette information est-elle interceptée ?

Les informations contenues dans l'en-tête HTTP des requêtes sont fournies par le navigateur au serveur Web auquel l'internaute se connecte. Si le site Internet consulté est un site dynamique, ses pages sont générées après réception de la requête du visiteur et avant envoi de la réponse du serveur. C'est appellation s'oppose aux sites Internet statiques dont les pages sont envoyées directement en réponse au client, sans subir de traitement préalable et donc sans aucune personnalisation possible. Le serveur décode alors l'en-tête HTTP de la requête du client pour en extraire les informations [30]. Celles-ci sont enregistrées dans des variables techniques appelées variables d'environnement CGI [31] (*Common Gateway Interface* ou interface de passerelle commune) dont le nom est préfixé de "HTTP" [32]. L'appellation CGI signifie que le serveur transmet la requête à un interpréteur propre au langage utilisé pour coder les pages du site Internet. L'interpréteur reçoit donc la requête, dont l'en-tête HTTP a été retiré, ainsi que les variables d'environnement du client et du serveur.

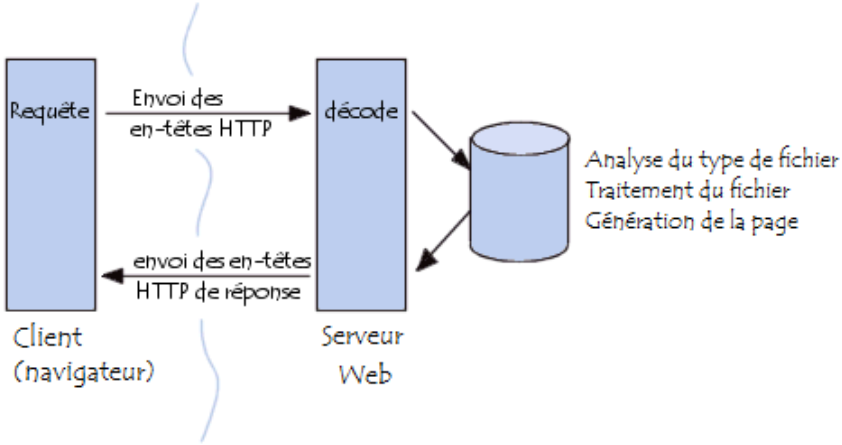


FIGURE 2.3 – Génération d'une page Web dynamique utilisant les variables d'environnement

Une fois la page interprétée, il génère un fichier qu'il transmet au serveur. Celui-ci ajoute un en-tête HTTP et envoie la réponse telle quelle au client, tel que schématisé à la figure 2.3. Ceci permet de créer des pages Web dynamiques, dont la présentation dépend de l'environnement du client et du serveur. Sans ce système d'interprétation, toutes les pages seraient statiques.

Quelle information est dévoilée ?

La liste des informations dévoilées et l'exploitation que peut en faire un tiers est trop longue pour être exposée de manière exhaustive. Les plus intrusives pour l'internaute étant les suivantes :

- Le lieu depuis lequel il se connecte, par géolocalisation de son adresse IP
- La langue de son navigateur, et donc très probablement sa propre langue
- Les sites internet visités précédemment durant la session de navigation
- Une liste de programmes installés, ceux permettant d'ouvrir les types de fichiers indiqués par le navigateur
- Les plugins installés sur le navigateur, qui sont parfois révélateurs de l'activité de l'internaute

Le cumul de ces informations forme une empreinte représentative du visiteur. Il y a tellement d'informations fournies que cette empreinte serait pratiquement unique dans le monde. Il s'agit donc là d'un autre moyen d'identifier un visiteur, qui pourrait tenter de se connecter de manière anonyme. Certaines des informations changement bien évidemment, telles que l'adresse IP mais

les autres données suffisent à reconnaître le visiteur. Ceci est d'autant plus vrai si la configuration de l'ordinateur du visiteur n'est pas courante. Par exemple, un internaute utilisant un système d'exploitation Linux et une résolution d'écran exotique sera très facile à reconnaître.

Afin de sensibiliser les internautes à ce sujet, et à titre d'exemple, l'EFF (*Electronic Frontier Foundation*), un organisme de défense des usagers d'Internet, a publié un projet de recherche établissant l'empreinte du navigateur et son caractère unique [3].

2.4 L'historique de navigation

L'activité sur internet et les sites consultés sont enregistrés à plusieurs niveaux.

Tout d'abord, l'historique du navigateur conserve la liste des sites internet avec les jours et heures de consultation. Il conserve également, pour un laps de temps limité, un enregistrement complet des pages et des images qu'elles contiennent. Ceci permet de recharger plus rapidement un site déjà consulté, puisque l'ordinateur possède déjà les informations demandées sur son disque [33]. Si cet historique a pour vocation de simplifier la tâche des utilisateurs en leur permettant de retrouver leur activité passée, il est à noter que cette information est accessible à toute personne utilisant l'ordinateur.

Dans le cas d'un réseau géré par un administrateur, l'infrastructure pourrait enregistrer tout le trafic transitant sur le réseau vers Internet. Une entreprise se voit même soumise à ce procédé par obligation légale, comme expliqué au chapitre précédent. Ce type d'enregistrement permet de retracer toutes les requêtes effectuées par chaque personne ainsi que les informations concernant les échanges de messages électroniques.

Les fournisseurs d'accès à Internet sont tenus légalement d'enregistrer, pour chacun de leurs clients, leurs informations personnelles récoltées lors de l'inscription au service, dont le nom, le prénom, la date de naissance, le sexe, l'adresse postale, le numéro de téléphone, les données bancaires ainsi que les périodes de paiement. De plus, de la même manière que les entreprises, ils se doivent de collecter les informations spécifiques à chaque session de navigation : l'adresse IP allouée, les sites visités, la nature de chaque opération, c'est-à-dire s'il s'agit de l'envoi d'un message électronique, du téléchargement d'une vidéo, etc. Bientôt, les FAI devront également enregistrer les métadonnées relatives aux messages électroniques échangés (voir section actualité). Concrètement, il ne s'agira pas de sauvegarder le contenu des messages mais les adresses d'expéditeur et de destinataires ainsi que les dates d'envoi ou de réception.

2.5 Les moteurs de recherche

Il arrive souvent qu'Internet ne soit pas utilisé pour la consultation mais pour la recherche d'information. C'est là que les moteurs de recherche entrent en jeu. Des centaines de serveurs scannent les différentes pages des sites Internet et, selon de savants calculs, les répertorient par catégorie selon des mots clés et par pertinence selon divers critères tels que les références depuis et vers d'autres sites pertinents, le nombre de visites, la propreté du code de la page, etc [34].

Certaines offres ont beaucoup évolué ces dernières années et sont passées d'annuaires en ligne à de véritables moteurs de recherche. Citons par exemple Google, Yahoo ou Bing. Les moteurs de recherche sont à présent le centre d'Internet, la porte d'entrée d'une session de navigation. Souvent, tout en sachant exactement où il souhaite se rendre, l'internaute préfèrera effectuer une

3. Cette expérience est accessible en ligne à l'adresse https ://panopticlick.eff.org

recherche sur un moteur plutôt que d'accéder directement au site souhaité, par simplicité. La recherche comporte plusieurs avantages. Elle pardonne les fautes de frappe ou d'orthographe. Elle accepte l'utilisation de raccourcis ("fb" recherché sur Google donnera Facebook comme résultat). Elle permet, pour certains sites, l'accès direct à une section plutôt que de passer par la page d'accueil du site et sélectionner la section souhaitée dans le menu. Finalement, il n'est nul besoin de retenir le nom exact, potentiellement long ou compliqué, des sites consultés, une recherche nous permet de les retrouver rapidement.

Par exemple, pour accéder au site de l'université de Namur, il suffit de taper "Université de Namur" dans la barre de recherche ou encore d'utiliser le diminutif "unamur" plutôt que l'adresse complète du site (www.unamur.be). De plus, les différentes sections du site de l'université de Namur sont proposées et accessibles directement via Google, comme on peut le constater à la figure 2.4.

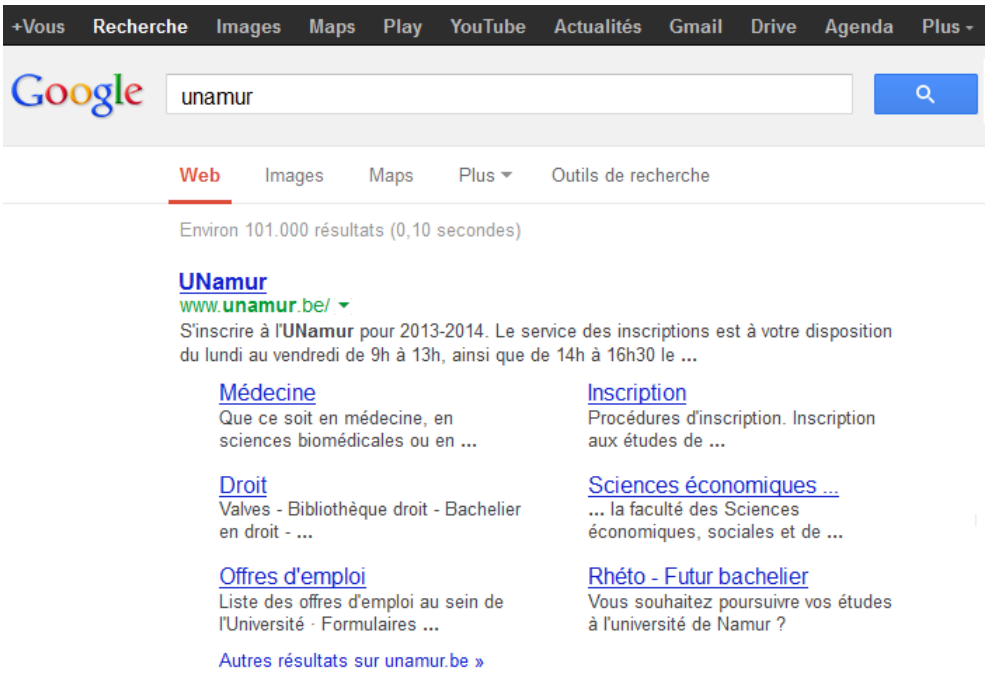


FIGURE 2.4 – Résultats de la recherche effectuée sur Google avec le mot clé "unamur"

Comment cette information est-elle interceptée ?

Le simple fait d'accéder au moteur de recherche donne déjà des informations au serveur. Il sait que l'internaute démarre une session de navigation. Si celui-ci se sert du moteur de recherche pour sa navigation, ce dernier sait, sans surprise, ce que recherche l'utilisateur et ne se prive pas d'enregistrer ces informations et de s'en servir pour le profilage. Ces données sont rendues anonymes au bout de neuf mois.

Quelle information est dévoilée ?

Le principal avantage d'un moteur de recherche se trouve également être son principal défaut du point de vue de la vie privée. En tant que point central d'Internet, et souvent point de départ d'une session de navigation, le serveur est au courant de la quasi-totalité des faits et gestes des internautes. Pire encore, s'il est défini en tant que page d'accueil, le serveur connaît

toutes les heures de connexion des internautes, même s'ils n'utilisent pas son service et naviguent directement vers un autre site. S'il est utilisé comme facilité pour atteindre d'autres sites, le moteur de recherche sait également quels sites l'internaute a pour habitude de consulter. Cela peut en dire long sur la vie privée de celui-ci, comme ses opinions politiques, son appartenance religieuse, ses centres d'intérêt, etc. Finalement, s'il est utilisé pour la fonction pour laquelle il a été conçu, le moteur de recherche connaît également tous les mots clés entrés par l'internaute. Ceci lui facilite la tâche, car il n'a plus besoin de regrouper les sites consultés dans des catégories, le visiteur lui fournit lui-même les mots clés.

2.6 Les communications électroniques

Nous l'avons vu dans la mise en contexte de ce document, sur Internet, les communications sont espionnées et enregistrées. C'est du moins le cas sur la plupart des services offerts gratuitement tels que Gmail, Outlook (anciennement Hotmail), Yahoo ! mail, Facebook, ou les divers services de création de blogs.

Le nombre de messages électroniques échangés chaque jour dans le monde s'élevait en 2010 à 294 milliards, ce qui en fait le mode de communication le plus utilisé de nos jours [35]. Chaque message électronique envoyé depuis ou vers une boîte mail gratuite est potentiellement scanné à la recherche de mots-clés et les pièces jointes sont analysées. Le fournisseur du service connaît donc la totalité des contacts, la fréquence des échanges de messages et surtout la nature de ces échanges pour chacun de ses clients.

En 2010, un porte-parole de la firme Google déclarait que "Gmail - comme la plupart des fournisseurs de webmail - utilise un scan automatique pour combattre le spam et les virus. Nous utilisons une technologie similaire pour afficher des publicités qui aident à garder nos services gratuits. C'est de cette façon que Gmail a toujours fonctionné." [36]. Les règles de confidentialité du service précisent que "L'utilisation de Gmail n'enfreint pas la confidentialité des émetteurs dans la mesure où nul autre que le destinataire est autorisé à lire le contenu des emails, et nul autre que le destinataire voit les annonces ciblées et des informations connexes." [37]. Or, nous avons pu constater dans la section traitant de l'actualité que les mails échangés ne sont pas seulement scannés, mais sont également conservés, puisque l'entreprise américaine a pu transmettre des échanges de mails à l'Etat belge notamment. Google est cité à titre d'exemple mais la collecte, l'analyse et l'exploitation des données échangées est bien évidemment du fait de nombreux services de messagerie en ligne, Microsoft et Yahoo ! appliquant les mêmes pratiques.

Les programmes de messagerie installés sur l'ordinateur, tels que le célèbre MSN Messenger, enregistrent toutes les conversations sur le disque dur de l'ordinateur. Si ces informations ne sont pas exposées sur Internet, elles sont tout du moins accessibles à toute personne utilisant l'ordinateur.

2.7 Les services de localisation

Il existe, sur Internet, des sites offrant des services de localisation d'adresse ou de consultation d'itinéraire. Lorsqu'ils sont utilisés, ces services retiennent les adresses recherchées par l'internaute. Pour ne citer que le plus célèbre, Google Maps propose même à ses utilisateurs d'enregistrer une série d'adresses prédéfinies telles que les lieux dits "maison" ou "bureau" en plus de l'enregistrement d'adresses dites favorites. Ces enregistrements nécessitent que l'utilisateur soit connecté au service via son compte personnel. Chaque information transmise ou recherchée est donc directement liée à son compte.

L'espionnage par localisation se répand avec la progression des terminaux mobiles. L'utilisation d'une application de cartographie ou de positionnement par GPS permet de connaître la localisation de l'utilisateur et même la signification de ce lieu. Comme c'est souvent le cas, les informations n'expriment toute leur valeur que lorsqu'elles sont croisées avec des informations récoltées par d'autres moyens. L'enregistrement de l'adresse de son domicile, comme mentionné précédemment, ou celui de l'adresse de ses contacts sont autant d'informations exploitables sur un *smartphone* ou une tablette.

D'autres techniques permettent encore de localiser l'utilisateur avec précision. Par exemple, si la carte WI-FI du terminal est allumée, elle peut scanner tous les réseaux environnants et, suivant les connexions disponibles à proximité et la puissance des signaux, déterminer à quelques mètres près, la position du terminal [38]. Encore une fois, ces informations ne sont exploitables qu'à la condition de connaître la position géographique des réseaux détectés. La société Google s'en est assurée en scannant un maximum de réseaux sans fils lorsque ses employés ont parcouru les rues du pays à bord des fameuses Google Cars, dont le seul but avoué était de photographier les rues en vue de préparer leur future application permettant de se promener virtuellement sur les routes (voir section sur l'actualité).

Combinées, ces techniques de localisation permettent de connaître à tout moment la position d'un internaute, pour palier le problème de la mobilité, rendant la géolocalisation par adresse IP inutilisable. Elles renseignent également sur les trajets effectués régulièrement par l'utilisateur, ses temps de trajet, ses heures de déplacement et les endroits présentant un intérêt.

2.8 Les formulaires en ligne

La manière la plus simple, pour un site, d'obtenir les informations d'un internaute reste de les lui demander lors de son inscription. Cela peut paraître anodin mais peu de sites peuvent garantir la confidentialité et la protection des données qui leur sont confiées. Si la protection s'avère insuffisante, les données privées peuvent être récupérées par une tierce personne. Egalement, si le gestionnaire du site est mal intentionné, il pourrait revendre les données de ses membres à des enseignes publicitaires. L'internaute serait alors surpris de recevoir des publicités mais n'aurait aucune idée de la manière dont ces entreprises ont obtenu son adresse.

Les données privées dont nous parlons ne sont dès lors plus du profilage anonyme mais touchent à l'identité civile de l'internaute. Nous parlons ici des nom, prénom, nationalité, langue, date de naissance, adresse postale, numéro de téléphone, adresse électronique, etc qui sont des champs typiques d'un formulaire et que l'internaute pourrait compléter naïvement s'il n'y prend pas garde. Les répercussions d'une telle fuite d'information sortent du seul cadre d'Internet et peuvent mener à bien des désagréments, tels que du démarchage téléphonique ou par voie postale.

2.9 Le Web social

Les blogs, les forums, les pages personnelles, les albums photos en ligne ou plus récemment les réseaux sociaux Facebook, Twitter et LinkedIn sont autant de composantes du web social, qui fait partie intégrante d'Internet depuis de nombreuses années.

Les communications sur les blogs et sur Facebook concernent la vie privée au regard des autres internautes car les informations ne sont plus uniquement scannées par des robots mais sont consultées par des êtres humains qui se connaissent éventuellement et peuvent avoir accès à

des renseignements que la personne concernée ne souhaite pas divulguer. Il n'est plus uniquement question d'exposer ses informations personnelles sur ce genre de services, mais également celle des autres personnes.

Les informations divulguées par ces moyens ne relèvent pas de la diffusion involontaire ou encore sous-jacente de données, car elles sont communiquées explicitement par une personne. Tout individu est considéré comme responsable et seul gérant de sa vie privée, ce qui peut être délicat pour les plus jeunes internautes.

Se pose donc ici un problème à part, celui du contrôle des informations par les individus, qui peuvent voir leur intimité violée par la publication d'une autre personne. Aucune action n'est alors possible pour remédier au préjudice, car le retrait des informations du service ne changement rien au fait que son public a d'ores et déjà pu prendre connaissance de l'information, et des copies peuvent en avoir été faites, manuellement ou par exemple par l'intermédiaire de notifications par courrier électronique.

Les médias sociaux sont de plus en plus régulièrement exploités par des professionnels dans l'espoir de glaner des informations concernant un client ou un collaborateur. Par exemple, les sociétés d'assurance pourraient se renseigner sur le risque engendré par un futur client, et décider, en fonction des résultats, de l'acceptation de couvrir ce risque. Un employeur pourrait entreprendre une démarche identique dans le but de préciser son opinion vis-à-vis des candidats à un emploi.

2.10 Le commerce électronique

Quel moyen plus efficace pour cibler la publicité que de surveiller les recherches et achats effectués dans les magasins en ligne ?

Ce type de commerce prend véritablement le pas sur les magasins, en se taillant une part de marché croissante chaque année [39]. Un rapport d'information du Sénat français fait établit le constat que "Les achats électroniques de biens et services par les particuliers sont estimés à 37 milliards d'euros en 2011, dont sept pour les courses de Noël. Ils ont triplé en 5 ans, si bien que la France se détache de la moyenne européenne pour se rapprocher de l'Allemagne, voire du Royaume-Uni. La Délégation à la prospective s'est interrogée, à un horizon de 10 ans, sur la durabilité et les conséquences de la croissance d'un secteur si bien-portant malgré la crise." [40].

Que ce soient de grandes plateformes de vente telles qu'eBay ou Amazon ou de plus modestes enseignes, les recherches, les achats et les ventes effectués sont souvent enregistrés et traités de manière à enrichir le profil de l'internaute. De façon assez évidente, les objets des recherches seront proposés dans de multiples bannières publicitaires, et les objets achetés ou vendus seront comparés à des bases de connaissance pour proposer l'achat d'autres biens en rapport avec ceux-ci.

Cette collecte de l'information est d'autant plus efficace qu'elle ne se base pas sur l'ensemble des activités de l'internaute pour proposer des produits qui, finalement, ont peu de chance de l'intéresser. Elle se base ici sur les activités menées sur des sites marchands et a donc une probabilité beaucoup plus importante de toucher le visiteur dans ses centres d'intérêt. Par exemple, l'achat d'une gamelle pour chien engendrera l'affichage de publicités pour accessoires canins, la recherche d'un appareil photo mènera les publicitaires à proposer des sites où ce modèle d'appareil est vendu, etc.

Il est une technique d'optimisation du chiffre d'affaire qu'il est intéressant d'analyser. Il s'agit du *yield management* ou "gestion fine". Elle consiste en l'adaptation systématique des prix en fonction du niveau de la demande. L'objectif est de maximiser le profit d'une entreprise. Cette méthode est particulièrement utilisée dans les secteurs des groupes hôteliers ou encore des transports de voyageurs où les frais sont fixes et où il convient de limiter le nombre de places vides, le coût d'un trajet étant quasiment fixe par rapport au nombre de voyageurs.

Le *Yield Management* est par exemple utilisé par la SNCF avec ses offres "Prem's", promotions de dernières minutes, tarifs plus élevés en période de pointe, etc. L'axe d'ajustement entre l'offre et la demande est donc le prix. Celui-ci permet de limiter les sous et sur-capacités de l'offre. Cet ajustement des prix amène l'entreprise à proposer ses services à des prix différents en fonction des personnes à qui elle s'adresse, de la période, et de ces propres capacités, le secret étant d'adapter ses offres au plus près des attentes de chaque segment du marché.

Les informations contenues dans les cookies ou l'adresse IP de l'utilisateur peuvent également être exploitées à ce genre de fins commerciales. La fréquentation et l'utilisation d'un site peuvent entraîner de rapides variations de prix. Si l'achat n'est pas immédiat, le site Internet peut conserver l'intérêt que l'internaute a manifesté pour une certaine offre dans le but de lui proposer un prix légèrement supérieur lors de sa prochaine visite, afin de susciter l'achat directe. Lorsque ce dernier retourne sur le site, à l'aide du même terminal, il constate la hausse du prix et, pensant que cela est dû à la diminution des stocks disponibles, sera incité à effectuer son achat sans délai. Et ceci même si le dis stock disponible est resté identique. Le prix peut ainsi augmenter plusieurs fois, par petits paliers, pour ne pas attirer l'attention lors de chaque simulation successive.

De même, les prix peuvent également varier en fonction de l'heure, celle-ci étant un bon indicateur du profil de l'acheteur. En effet, les achats de billets de train ou d'avion en semaine et durant les heures de travail sont typiquement réalisés par des entreprises pour des déplacements professionnels et qui sont généralement moins sensibles aux prix que les particuliers, d'où une tarification plus élevée. On pourrait aussi imaginer des algorithmes plus intelligents qui tiendraient compte d'autres informations tel que le nombre de passagers de la réservation. Par exemple, pour un même trajet, les particuliers sont certainement prêts à débourser un petit peu plus pour que toute la famille voyage dans le même appareil. Le système viserait alors à maximiser la ventes de ces extra.

2.11 Le *Cloud computing*

Une nouvelle forme d'informatique est née, suite à l'augmentation des vitesses de connexion à Internet. Il s'agit du *cloud computing* ou informatique dans les nuages, qui consiste à fournir un service en ligne où les données et paramètres sont enregistrés non plus sur le disque dur de la machine qui exploite le logiciel, mais sur un serveur distant, qui a la charge de ces données [41]. Ce type de service est proposé aux entreprises et aux particuliers, qui y trouvent de nombreux avantages tels que l'assurance de ne perdre aucune donnée, de pouvoir y accéder partout, depuis divers appareils (*smartphone*, tablette, autre ordinateur) ainsi que leur disponibilité constante, cela pour un coût réduit, surtout pour les entreprises qui peuvent diminuer voire supprimer leur parc informatique et donc les coûts d'achat, de maintenance et de personnel qui en découlent [42].

Il y a cependant un inconvénient majeur à une telle gestion des données, à savoir la perte de leur contrôle, étant donné qu'elles sont confiées à un sous-traitant. Ce dernier doit assurer la sécurité d'accès à ces données mais ne tombe pas nécessairement sous la même législation que ses clients. Les données peuvent ainsi être enregistrées physiquement n'importe où dans le monde, et sortir du cadre juridique européen [43]. Citons le cas du *Patriot Act* aux Etats-Unis,

voté en 2001, qui autorise l'Etat américain à accéder à toutes les données des entreprises et des particuliers sur simple demande, à partir du moment où la société offrant le service en ligne est basée aux Etats-Unis ou y a des intérêts économiques. Dans ce cas précis, l'entreprise ne peut assurer la confidentialité des informations qui lui sont confiées, même si celles-ci ne quittent pas le territoire européen.

A titre d'exemple pour les utilisateurs particuliers, nous pouvons regrouper les réseaux sociaux, les blogs, les messageries électroniques en ligne, les sites d'achat et bien plus encore, qui sont pour la grande majorité fournis par des sociétés américaines. Les options de sauvegarde en ligne des données et paramètres sont activées automatiquement sur le navigateur Chrome, lorsque l'utilisateur est connecté à l'aide de son compte. L'historique de navigation, les identifiants et mots de passe, les saisies de formulaire, les favoris, tout est enregistré automatiquement sur les serveurs de Google. L'avantage mis en valeur étant de retrouver toutes ces informations simplement en se connectant à l'aide de son compte sur une autre machine, mais l'inconvénient étant bien entendu de fournir toujours plus d'informations sur son activité, allant ici jusqu'aux mots de passe.
Un service identique est proposé par Mozilla, via l'outil Sync de son navigateur Firefox. Une différence de taille toutefois étant que ce dernier propose l'usage de ce service à l'utilisateur, et ne l'exploite pas sans l'en informer, comme son concurrent.

2.12 Les informations sur les mobiles

Cette section, un petit peu particulière, traite de problèmes de divulgation d'informations propres aux appareils mobiles. Cest derniers étant de nos jours largement répandus, il est indispensable d'en survoler les aspects.

Le problème de la mobilité, qui empêche la géolocalisation par adresse IP, est résolu, chez Google, par des demandes insistantes d'associer son numéro de téléphone mobile au compte client, sous couvert de protection des données de l'utilisateur et de récupération de mot de passe, si ce dernier venait à l'oublier. Ainsi, la firme lie les activités de navigation mobile aux autres activités de l'utilisateur. L'utilisation du système pour *smartphone* Android requiert l'enregistrement du compte Google sur le téléphone. La boucle est bouclée, toute activité effectuée par l'internaute, que ce soit sur son ordinateur ou sur son téléphone, est enregistrée dans un unique profil des plus complets, contenant des informations aussi variées que les heures de connexion, les recherches effectuées sur Internet, les messages électroniques échangés, les réseaux sans fil utilisés ainsi que leurs mots de passe, les favoris enregistrés dans le navigateur, le numéro de téléphone de l'utilisateur, ses contacts et leurs numéros de téléphone respectifs, tous les détails concernant les déplacements effectuées, la position actuelle, et éventuellement d'autres informations dont nous ne soupçonnons même pas l'enregistrement.

Un comportement similaire est remarqué chez Microsoft qui, allant plus loin que son concurrent, oblige les clients de sa messagerie Outlook à indiquer une messagerie de secours et, s'ils n'en possèdent pas, leur numéro de téléphone, précisant par la même que si l'information n'est pas fournie dans les jours suivant la requête, le compte de messagerie électronique de l'internaute sera bloqué jusqu'à régularisation de son profil [44].

Lors de l'utilisation d'un smartphone, il est proposé à l'utilisateur d'activer des options de géolocalisation. Ces options sont présentées comme des améliorations du comportement de certaines applications, sans en préciser les détails, mise à part pour les applications de navigation, où il est indiqué que l'activation de la carte réseau sans fil permet d'accélérer la localisation de l'utilisateur, par rapport à la seule utilisation de la puce GPS. Néanmoins, lorsqu'on y prête

attention, de nombreuses applications demandent le droit d'accès à la position géographique de l'utilisateur sans en préciser la raison. Si l'utilisateur a activé les options de géolocalisation sur son appareil, ces informations sont envoyées de manière invisible pour celui-ci, toute application peut donc potentiellement accéder aux coordonnées de l'utilisateur, en temps réel.

Une autre menace d'exposition de données privées survient lorsque l'utilisateur télécharge et installe une nouvelle application pour son appareil mobile. Il est prévu, par la plateforme d'hébergement des applications, que chacune d'elles indique les informations auxquelles elle nécessite l'accès pour pouvoir fonctionner. Il arrive cependant très régulièrement que des applications indiquent requérir l'accès à une ressource du téléphone qui n'est pas en rapport avec son activité. Par exemple, une application lampe de poche peut demander les droits d'accès aux messages, à la liste des contacts ou encore à l'agenda, enregistrés sur l'appareil.

2.13 Les logiciels espions

Certains logiciels, installés consciemment ou non, espionnent l'internaute encore bien au-delà de ce qui a été énoncé jusqu'à présent. Cela peut aller de la surveillance des habitudes de navigation jusqu'à la récupération des mots de passe et autres informations sensibles telles les informations bancaires. Il existe deux approches pour ces logiciels, à savoir s'installer et s'exécuter de manière silencieuse, sans que l'utilisateur ne puisse les repérer, ou s'installer et s'exécuter au grand jour, sous prétexte de fournir un ou plusieurs autres services à l'usage de l'utilisateur.

Ceux utilisant la première approche sont cachés sous couvert de fichiers pour lesquels l'internaute porte un intérêt et, une fois téléchargés, deviennent actifs sur l'ordinateur. Ils transmettent dès lors toutes sortes d'informations au sujet de l'internaute et peuvent découvrir les mots de passe et comptes bancaires utilisés sur l'ordinateur infecté, notamment par la technique du *keylogging*, qui consiste à enregistrer toutes les frappes entrées au clavier et à les transmettre à l'auteur du logiciel espion. Il lui est alors aisé de trier ces données pour en ressortir des informations de valeur telles que des adresses e-mail, des noms et prénoms, des adresses, des numéros de carte de crédit, etc.

Parmi ces logiciels, il en est un à surveiller car son rôle reste pour le moment un mystère. Le navigateur Google Chrome active, lors de son installation, une librairie s'exécutant automatiquement sur l'ordinateur. Cette librairie reste présente après désinstallation du navigateur, et agit donc de manière indépendante. Il semblerait qu'elle espionne les activités menées à l'aide du navigateur concurrent Internet Explorer [45].

D'autres logiciels utilisent la deuxième approche et échappent à la surveillance des logiciels anti-virus, du fait de leur légitimité apparente. Les plus courants sont les barres de recherche intégrées au navigateur. Ces petits utilitaires sont souvent proposés par défaut lors de l'installation d'autres programmes et sont installés si l'utilisateur n'y prend pas garde. De nombreuses personnes remarquent ainsi l'apparition d'une ou plusieurs barres de recherche dans leur navigateur sans en connaître la provenance ni l'utilité et ignorent la marche à suivre pour s'en débarrasser, ni même si cela est possible. De nombreuses sociétés ont développé leur propre barre de recherche, qui permettraient à l'internaute d'effectuer des recherches ciblées, voire d'effectuer une navigation sur Internet soit disant plus sûre. Ces barres, inclues au navigateur, ont accès à l'entièreté des sessions de navigation de l'internaute, puisque présentes sur toutes les pages qu'il visite.

Nombre des barres de recherche détournent la page de démarrage du navigateur, pour la supplanter à l'aide d'une page imitant un moteur de recherche mais qui ajoute des publicités au sein de la liste de résultats tirés du moteur de recherche exploité.

2.14 Récapitulatif des techniques de collecte et d'exploitation d'informations

Après avoir explicité les différents moyens les plus courants d'obtention des données relatives aux internautes et leurs exploitations possibles dans le cadre du profilage, et avant d'aborder les méthodes de protection, établissons un tableau comparant les informations potentiellement recueillies pour chacun de ces moyens. L'acquisition de ces informations n'étant pas toujours une finalité en soi mais permettant d'en obtenir d'autres, plus utiles ou plus personnelles, elles ont été regroupées en catégories logiques en fonction des données auxquelles celles-ci donnent accès :

- **Identité réelle** : Les informations légales relatives à l'internaute, permettant de l'identifier, telles que définies dans la loi "vie privée".
- **Identité fictive** : Toute information permettant de reconnaître l'internaute entre deux sessions de navigation, et d'établir un profil le concernant, sans pour autant connaître son identité réelle.
- **Données sensibles** : Toute information strictement confidentielle pour l'internaute, tels que les mots de passe et les numéros de carte de crédit.
- **Données personnelles** : Informations relatives à l'identité réelle de l'internaute mais ne permettant pas de déduire cette identité ni d'effectuer une reconnaissance entre deux connexions (âge, sexe, langue, etc).
- **Localisation** : La position géographique approximative de l'internaute, voire son adresse postale.
- **Centres d'intérêt** : Les thèmes qui tiennent à coeur à l'internaute et pourront être exploités principalement dans le cadre de publicités ciblées.
- **Pages visitées** : La liste complète ou une partie des sites Internet visités avec l'adresse, la date, l'heure, etc.
- **Matériel utilisé** : Reconnaissance de l'environnement de connexion de l'internaute, comme son ordinateur, la version du système d'exploitation, le type de navigateur, etc.

	Identité réelle	Identité fictive	Données sensibles	Données personnelles	Localisation	Centres d'intérêt	Pages visitées	Matériel utilisé
L'adresse IP	■				■			
Les cookies		■					■	
Les en-têtes HTTP		■		■			■	■
L'historique de navigation							■	
Les moteurs de recherche						■	■	
Les communications électroniques	■			■	■	■		
Les services de localisation					■			
Les formulaires en ligne	■			■	■	■		
Le Web social	■			■	■	■		
Les commerces électroniques	■		■	■	■	■		
Le *Cloud computing*		■		■		■		
Les informations sur les mobiles		■		■	■	■	■	■
Les logiciels espions	■	■	■	■	■	■	■	■

Chapitre 3

Analyse des techniques de défense de l'utilisateur

Lors de l'analyse des diverses techniques de collecte d'informations ainsi que des traces laissées sur Internet lors des sessions de navigation, nous avons pu constater la grande quantité et diversité des données disséminées sur le parcours des internautes. S'il n'est pas possible de se rendre anonyme sur la toile, il existe des outils tout aussi variés pour limiter les traces laissées dans son sillage.

Les technologies de protection des données personnelles, aussi appelées les PETS (*Privacy-Enhancing Technilogies*) désignent les outils visant à encadrer et réduire, autant que possible, la collecte des données à caractère personnel [46]. La Commission européenne et les gouvernements nationaux tentent de promouvoir ces outils auprès des particuliers car ils constituent pour le moment les seuls moyens de prévention à leur disposition. "Les gouvernements ont perdu le contrôle sur la technologie, la normalisation des terminaux et protocoles de communication sont définies par des organisations privées ou des consortium d'entreprises commerciales, la législation privée se concentre sur les responsables de traitement et ne régule pas la technologie en tant que telle..." [1].

Dans ce chapitre seront expliquées des techniques accessibles à tout un chacun pour protéger ses données personnelles, de la plus simple à la plus complexe, de l'intervention unique à l'action répétitive. Si certaine sont largement répandues et réputées pour leur efficacité, d'autres sont moins évidentes ou plus subtiles et permettent de compléter les premières. L'objectif de ce chapitre est d'élargir au maximum le champ des informations protégées afin de couvrir toutes les fuites exposées au chapitre précédent.

Tour à tour, chaque technique sera détaillée puis diverses caractéristiques seront analysées afin de pouvoir les comparer. Chaque technique se verra ainsi attribuer une note allant de 1 (médiocre) à 5 (excellent) pour chacune des caractéristiques considérées. Afin d'évaluer correctement ces techniques, elles ont été testées tour à tour pour se faire une idée précise des processus d'installation et d'utilisation. Leur efficacité a ensuite été mise à l'épreuve dans la mesure du possible. Les notes attribuées aux caractéristiques d'une technique ne sont pas absolues, elles sont relatives aux performances observées par les autres techniques. C'est pourquoi la cotation a été pondérée afin de refléter les avantages ou inconvénients majeurs que chaque technique offre par rapport aux autres. Par exemple, une méthode de protection nécessitant une heure de recherches et de comparaisons des offres sera mal cotée, puisque la plupart des solutions abordées dans ce

1. Déclaration de Jean-Marc Dinant, directeur de l'unité "technologie et sécurité" au Centre de Recherche Informatique et Droit (CRID) de l'Université de Namur à propos de la vie privée sur Internet.

document sont applicables en quelques minutes.

Les critères retenus pour qualifier une technique sont :

Installation : Simplicité du processus d'installation, simplicité de la solution ou technologie concernée, connaissances techniques indispensables à maîtriser, recherches à entreprendre ou encore temps nécessaire.

Répétition : Durée de validité de l'installation, action unique ou à réitérer lors de chaque utilisation, le processus d'installation doit-il être répété entièrement ou seulement en partie.

Utilisation : Avantages et contraintes engendrés par l'utilisation de la technique (ex : un service plus rapide ou au contraire plus lent, limitation des possibilités d'effectuer certaines actions, possibilité d'en exécuter de nouvelles).

Efficacité : Quantité et qualité des informations protégées, caractère unique de la solution. Une technique ayant l'exclusivité de la protection d'une information aura, si cette information a un impact important sur la protection de la vie privée de l'internaute, une note équivalente à une technique permettant de couvrir de nombreuses données moins vitales ou pouvant également être protégées via d'autres techniques.

Entrent aussi en ligne de compte, dans une moindre mesure :

Coût : Prix de mise en oeuvre et d'utilisation de la technique. Il est à remarquer encore une fois que ce critère se base sur la comparaison du coût de chaque technique. La plupart d'entre elles étant gratuites, une solution peu coûteuse sera tout de même fortement pénalisée. Une solution ayant des déclinaisons gratuites et payantes verra sa note baisser si la solution payante a de fortes chances d'être envisagée par l'internaute.

Portabilité : Adaptabilité de la solution à divers environnements, c'est-à-dire différents terminaux (ordinateur, *smartphone* ou tablette), différents systèmes d'exploitation, différents navigateurs, etc.

Les techniques abordées suivront toutes le principe suivant :

1. Si cela est possible, cacher totalement les informations.
2. Sinon, limiter leur impact :
   a) Cacher ce qui est possible.
   b) Répartir les informations entre plusieurs services pour limiter leurs connaissances respectives de l'utilisateur.
   c) Obscurcir les informations en partageant le service avec d'autres utilisateurs afin de déjouer les algorithmes d'espionnage.

3.1 La déconnexion systématique

Tout comme on éteint la lumière en changeant de pièce, il est de bonne habitude de se déconnecter d'un service lorsque l'on en a plus l'usage. Ceci permet d'éviter que ce service reste à l'écoute des actions effectuées par l'utilisateur. Ceci est particulièrement valable dans le cas des entreprises fournissant de multiples services sur Internet, donc certains nécessitent d'être authentifiés, tandis que les autres sont libres d'utilisation.

Citons par exemple, les services de messagerie électronique gratuits, qui nécessitent naturellement une authentification. Après avoir consulté ses messages, si l'internaute ne se déconnecte pas, il reste identifié auprès de la société offrant le service. S'il exploite un autre service, il sera

dès lors suivi sans aucune difficulté. Il peut par exemple utiliser un moteur de recherche, consulter des vidéos, ou toute autre activité sans avoir conscience d'être toujours connecté à son compte.

Il en va de même pour tous les services requérant une authentification tels que les services de stockage en ligne, les services de partage de photos et de vidéos, les réseaux sociaux, etc. Toute technique de protection de la vie privée est rendue caduque si l'internaute indique lui-même son identité au service qu'il exploite.

Quels sont les avantages ?

Les sociétés de service collectent les informations relatives à l'utilisation qui est faite de leurs services, il est inutile de leur simplifier davantage la tâche en indiquant ouvertement son identité. La déconnexion des comptes utilisateurs est un prérequis à l'utilisation des techniques qui seront abordées par la suite.

Ce simple geste participe à renforcer la sécurité, puisqu'une tierce personne ne pourra plus usurper l'identité de l'utilisateur en profitant de ses comptes restés connectés. Ceci est particulièrement valable sur un ordinateur dont l'accès est public.

Quels sont les inconvénients ?

Avant chaque utilisation d'un service, l'utilisateur doit se connecter, ce qui constitue une étape supplémentaire et un léger désagrément. Il en va de même après l'utilisation du service, où l'utilisateur doit se rappeler systématiquement de se déconnecter de son compte client.

Il s'agit là de deux actions répétitives qui peuvent paraître fastidieuses, surtout si elles sont effectuées plusieurs fois par jour, mais c'est une bonne habitude à prendre, qui rendra rapidement ce geste automatique.

Tableau d'évaluation

Critère	Note	Justification
Installation	★★★★★	Rien ne doit être installé ou configuré.
Répétition	★☆☆☆☆	Se connecter à chaque utilisation.
Utilisation	★★☆☆☆	Etapes de connexion et déconnexion supplémentaires et non automatiques.
Efficacité	★★★☆☆	Cache les informations d'utilisation des autres services.
Coût	★★★★★	Gratuit.
Portabilité	★★★★★	S'applique à tous les dispositifs.

3.2 Les moteurs de recherche anonymes

Certains moteurs de recherche ont été créés afin de garantir une meilleure protection de la vie privée de leurs utilisateurs en n'enregistrant pas leurs historiques de navigation. En particulier, les heures de navigation, les adresses IP, les recherches effectuées, les liens cliqués et les cookies ne sont pas utilisés, et ne sont pas non plus transmis à des sites tiers tels que des agences publicitaires.

C'est le cas de **DuckDuckGo**, un moteur de recherche qui met l'accent sur la recherche anonyme, chiffrée de bout en bout, et qui ne stocke aucune information personnelle ni de navigation [47]. Il ne transmet pas non plus d'informations aux sites utilisés pour compiler les pages de résultats. Le site n'enregistre pas les requêtes effectuées et ne permet pas à d'autres sites de le faire, s'opposant ainsi au profilage. Il favorise aussi l'utilisation d'HTTPS, protocole de communication sécurisé par chiffrement des informations échangées, sur les sites auxquels il renvoie dans les résultats de recherche. Depuis le scandale du projet PRISM (voir section sur l'actualité), le navigateur DuckDuckGo a enregistré une hausse de popularité importante, à savoir 77 millions de recherches effectuées en juin 2013, contre 54 millions le mois précédent [48].

Une autre alternative aux moteurs de recherche populaires est le métamoteur [2] de recherche **Ixquick**, lancé en 1998, ou le moteur de recherche **Startpage**, lancé en 2009 par la même société (Surfboard Holding BV). Tous deux offrent la garantie d'une navigation anonyme sans enregistrement d'aucune information concernant l'utilisateur [49]. Les communications ne sont cependant pas chiffrées.

DuckDuckGo et Ixquick soutraitent les recherches auprès de nombreux moteurs célèbres, puis compilent les résultats avant de les présenter à l'utilisateur. Startpage ne travaille qu'avec Google mais fournit des options de recherche avancées telles que la langue, la région, le type de fichier cherché, la date, etc.

Il est, d'une manière générale, recommandé d'utiliser plusieurs moteurs de recherche différents afin de ne pas laisser toutes ses traces au même endroit, suivant le principe de la protection des données privées.

Quels sont les avantages ?

Aucune information personnelle ou concernant les recherches effectuées n'est collectée, par qui que ce soit. L'utilisation du chiffrement assure que les informations ne peuvent pas non plus être récupérées de manière illicite.

Les résultats des recherches sont garantis authentiques, ils ne sont pas filtrés en fonction du profil de l'utilisateur. L'accès aux informations est donc total.

Quels sont les inconvénients ?

Les résultats de recherche ne sont pas filtrés en fonction du profil de l'utilisateur, il est donc possible qu'ils soient moins pertinents et que l'utilisateur doive fournir un effort supplémentaire pour trouver l'information recherchée.

L'historique de recherche est perdu pour l'utilisateur aussi, ce qui est un moindre mal puisque ce dernier peut encore compter sur l'historique de la saisie des champs enregistré par le navigateur,

2. Un métamoteur est un moteur de recherche qui puise ses informations à travers plusieurs moteurs de recherche généralistes.

ainsi que sur l'historique de navigation, qui fournit les pages Internet consultées, ce qui est souvent plus utile que de retrouver les mots clés utilisés.

Tableau d'évaluation

Critère	Note	Justification
Installation	★★★★★	Rien ne doit être installé.
Répétition	★★★★★	Peut être configuré comme moteur de recherche par défaut et/ou comme page d'accueil.
Utilisation	★★★★★	Aussi performant qu'un moteur de recherche classique (moins ciblé mais plus objectif).
Efficacité	★★★☆☆	Cache les sujets de recherche et l'historique de navigation.
Coût	★★★★★	Gratuit.
Portabilité	★★★★★	S'applique à tous les environnements.

3.3 Les identités virtuelles

Sur Internet, rien n'oblige l'internaute à divulguer sa véritable identité, tant que l'organisme qui la réclame n'est pas une institution légale. Les services de messageries électronique, les forums, les réseaux sociaux, et tout autre service privé ne peuvent forcer les internautes à révéler des informations qu'ils jugent privées et ne souhaitent pas communiquer. Il est dès lors judicieux de préparer des identités virtuelles qui pourront être utilisées lors de l'inscription à l'un de ces services. La clé de l'anonymat étant, lorsqu'on ne peut cacher ses informations, de les falsifier et surtout de les faire varier d'un site à l'autre, pour éviter que ces identités virtuelles soient suffisamment complètes pour permettre de remonter jusqu'à l'identité réelle de l'internaute. Il est important de créer ces identités virtuelles de toute pièce, car se baser sur les informations personnelles d'une connaissance reviendrait à usurper son identité et pourrait causer des torts tant à la personne concernée qu'à l'internaute.

Afin de gagner du temps et épargner à l'utilisateur l'effort de créer une identité virtuelle, des services permettent d'emprunter temporairement une identité prête à l'emploi. Les informations les plus régulièrement demandées pour accéder à un service sur Internet sont l'adresse électronique, et la création d'un compte lié au site fournissant le service. Ces situations consistent en une divulgation inutile des informations privées, surtout s'il s'agit d'utiliser un service une seule fois.

Les adresses électroniques fournies sont la plupart du temps vérifiées par l'envoi d'un message de confirmation que l'utilisateur doit consulter. Il est donc inutile de fournir au service une adresse inexistante. Pour pallier ce problème, il existe des services de messagerie électronique proposant des adresses dites jetables, tel que **Yopmail**. Ce service autorise l'utilisateur à choisir une adresse liée au nom de domaine de Yopmail, par exemple "adresse123@yopmail.com". Cette adresse peut dès lors être utilisée pour une inscription à un service, les messages reçus seront consultables à partir du site de Yopmail. Les adresses ainsi utilisées ne sont jamais créées ni

supprimées, elles n'existent pas. Lorsque le serveur de messagerie de Yopmail reçoit un courrier, il l'enregistre dans sa base de données en vis-à-vis de l'adresse à laquelle le message est destiné. Lorsqu'un utilisateur interroge le serveur à propos du contenu de la boîte de messagerie liée à l'adresse, ce dernier renvoie la liste des messages correspondant à cette adresse. Les messages reçus peuvent être supprimés manuellement, mais le seront automatiquement après quelques jours. Cette technique permet à l'internaute d'éviter que son adresse de courrier électronique ne soit reprise dans les bases de données de dizaines de sites, ce qui lui permet de contrôler véritablement sa distribution et ainsi éviter de recevoir quantité de messages publicitaires. Ce type d'adresse n'autorise pas l'expédition de messages, mais seulement leur réception.

Dans la même optique, et si l'utilisation d'adresses jetables devient courante, l'internaute peut créer une adresse de messagerie alternative qu'il fournira aux indésirables. De nombreux fournisseurs de service proposent également la création de multiples alias liés à l'adresse de messagerie. Ils ont alors le même comportement qu'une adresse séparée mais offrent l'avantage d'être tous consultables en quelques instants, puisque l'utilisateur ne devra pas se connecter au service. La connexion à son adresse de messagerie suffit à consulter cette dernière et tous les alias qui lui sont liés. Les alias sont créés et supprimés en quelques clics.

Suivant un principe identique, le site **Bugmenot** permet d'utiliser les services en ligne nécessitant la création d'un compte. Il propose à ses utilisateurs de partager leurs différents comptes, liés à des services sur Internet, pour que les autres utilisateurs puissent en profiter et éviter de s'inscrire eux aussi auprès du service exploité. L'internaute qui se sert de telles identités publiques doit veiller à ne fournir au service exploité aucune information qu'il ne désire pas partager car elle serait liée au compte public.

Quels sont les avantages ?

L'avantage principal de l'utilisation de ces stratagèmes est évidemment la conservation des informations personnelles de l'internaute, qui n'est plus contraint de les divulguer pour accéder à un service.

La création et l'utilisation d'identités virtuelles permet de fournir des informations cohérentes, choisies par l'utilisateur, et assure que ce dernier sera le seul à en faire usage. Les adresses électroniques jetables et les comptes publics offrent également un gain de temps considérable puisqu'ils évitent à leur utilisateur l'étape de création de comptes.

Quels sont les inconvénients ?

La création et la gestion de multiples identités virtuelles est fastidieuse, car l'internaute doit également se souvenir de l'adresse qu'il a utilisé pour s'inscrire à chaque service.

Les adresses électroniques jetables et les comptes publics ne peuvent être utilisés que dans le cadre d'une utilisation où aucune information personnelle ne sera fournie. Dans le cas contraire, ce ne serait plus uniquement le service exploité qui en aurait connaissance, mais également les internautes qui feront usage du compte public dans le futur.

Il est à noter que ces services ne conviennent pas pour toutes les situations. Les messageries électroniques jetables ne permettent pas d'envoyer de messages, mais seulement d'en recevoir. Puisqu'ils sont basés sur la contribution de la communauté, les comptes publics ne sont pas toujours disponibles pour chaque service sur Internet, il est nécessaire qu'un internaute crée un compte en utilisant une identité virtuelle, puis le partage sur le site de partage des comptes publics pour qu'il soit accessible au reste de la communauté.

Tableau d'évaluation

Critère	Note	Justification
Installation	★★☆☆☆	Les identités doivent être créées.
Répétition	★★★★★	Une fois créées, les identités restent valables.
Utilisation	★★★★☆	Gain de temps, pas de formulaire d'inscription à remplir.
Efficacité	★★★★☆	Le profilage basé sur une identité virtuelle.
Coût	★★★★★	Gratuit.
Portabilité	★★★★★	S'applique à tous les environnements.

3.4 Les proxys

Une communication typique entre deux appareils, qu'il s'agisse de deux ordinateurs ou d'un client et d'un serveur, passe par des intermédiaires. Il peut s'agir de routeur, de pare-feu, de proxy, etc. Certains ont un simple rôle d'acheminement des données, tandis que d'autres pratiquent une analyse, voire un traitement des données qui transitent.

Le proxy, qui signifie littéralement "mandataire", se place en tant qu'un de ces intermédiaires. Il reçoit les requêtes de ses machines client et se charge de les transmettre à la machine distante, en plaçant son propre en-tête HTTP. La machine distante, typiquement un serveur Web, répond alors au proxy, qui transmet cette réponse à son client qui a initié la transaction. Cet échange est illustré à la figure 3.1. Un proxy est un logiciel mais désigne souvent, par extension, la machine sur laquelle il est installé, celle-ci étant souvent dédiée entièrement à cette tâche.

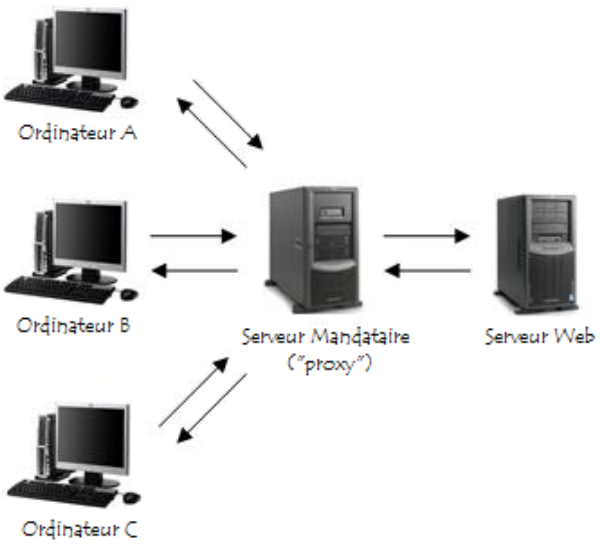


FIGURE 3.1 – Le proxy agit comme un intermédiaire entre deux machines

La position centrale du proxy au sein de la communication entre un client et un serveur lui

permet de jouer de nombreux rôles [50].

Le **proxy filtrant** contrôle le trafic qui le traverse et bloque l'accès à certains services. Cette utilisation rend l'usage de proxys courante dans les grandes infrastructures car ils facilitent le contrôle de l'utilisation qui est faite du réseau, et principalement d'Internet, en mettant en place une liste noire ou une liste blanche d'adresses IP et de noms de domaines. Le personnel d'une entreprise pourrait ainsi se voir refuser l'accès aux boîtes de messagerie privées, aux journaux en ligne, ou tout autre service sortant du cadre du travail dans l'entreprise.

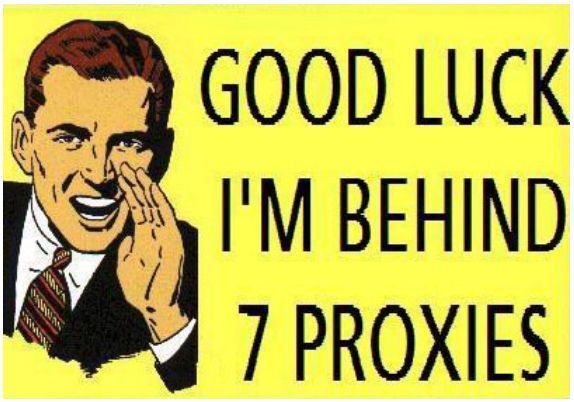
Le proxy peut jouer un rôle de **protection** du parc informatique interne en autorisant les connexions de l'intérieur vers l'extérieur, mais en refusant à des machines distantes, sur Internet, de se connecter à un ordinateur interne.

Le **proxy de contournement** permet d'outrepasser les filtrages mis en place pour restreindre l'accès à certains services. Une telle utilisation est exactement inverse à la première, les proxys filtrants. Le proxy est utilisé ici pour contourner les règles de filtrage d'un autre proxy. Le proxy filtrant auquel nous sommes soumis ne perçoit plus la machine distante avec laquelle nous communiquons, mais uniquement le proxy par lequel nous passons. L'accès au proxy de contournement doit évidemment être autorisé par le proxy filtrant à contourner, sans cela aucune connexion ne peut être établie. Ceci permet, par exemple, de consulter des sites bloqués par les autorités d'un pays ou encore par le propriétaire d'un site Internet.

Le **proxy anonymiseur** fournit un service qui découle directement du précédent, l'anonymisation passe par la prise d'identité du proxy. Le serveur qui reçoit les requêtes ne voit plus l'adresse de son communiquant, mais uniquement l'adresse du proxy qui est utilisé. Les informations qu'il pourra collecter seront donc erronées, ceci est d'autant plus vrai que plusieurs clients passeront par le même proxy. Les données seront alors mélangées sans pouvoir être distinguées l'une de l'autre. Cette technique s'appelle l'assombrissement (*obfuscation*) des données et revient à "se fondre dans la masse" puisque tous les internautes passant par le même proxy ont la même adresse IP. Cette technique a toutefois ses limites et ne peut garantir l'anonymat à elle seule. Par exemple, comme expliqué précédemment, les en-têtes HTTP incluent des informations tellement nombreuses qu'ils forment, une fois combinés, une empreinte identifiant une machine de manière quasiment unique. Un internaute exploitant un proxy mais utilisant une machine équipée de Linux avec une résolution d'écran exotique et utilisant un navigateur peu répandu sera immédiatement repéré. Il faut impérativement conserver une configuration très répandue ou exploiter les possibilités offertes par l'utilisation de machines virtuelles.

L'anonymisation par proxy est une technique fréquemment utilisée par les pirates informatiques pour pouvoir agir en toute impunité. La procédure est bien évidemment plus complexe et passe par une chaîne de proxys, de préférence répartis partout dans le monde, et peut même passer par le détournement de l'ordinateur d'un particulier, pour se servir de sa machine comme serveur proxy, à son insu. Retrouver l'identité du pirate nécessite dès lors de contacter les fournisseurs de chaque service proxy un à un pour remonter la chaîne. L'utilité de pirater l'ordinateur d'un particulier pour s'en servir prend dès lors tout son sens, n'étant pas un service organisé, les requêtes transitant sur cet ordinateur ne sont pas journalisées, et ne peuvent donc être retrouvées.

Le **proxy de temporisation** permet de limiter la quantité de données échangées effectivement sur un réseau. Chaque ressource demandée est téléchargée par le proxy, puis enregistrée sur un disque pour pouvoir être réutilisée en cas de requête ultérieure d'accès à la même ressource. Certains fournisseurs d'accès à Internet ou entreprises conséquentes utilisent cette technique pour diminuer la charge de leur réseau [51].



Le proxy peut donc ne pas se limiter au simple rôle de passerelle. Un exemple célèbre de programme mandataire est **Privoxy** dont le nom est la concaténation de "privacy" et de "proxy". Il propose à ses utilisateurs des options avancées de protection de la vie privée en filtrant les pages Web consultées, gérant les cookies envoyés par les serveurs Web, contrôlant les accès entrants vers l'ordinateur, supprimant les publicités et autres fenêtres intempestives appelées pop-up.

Les tunnels sont des proxys chiffrés. Les données y sont cryptées, ce qui est encore plus sécurisé mais ce service est souvent payant.

Quels sont les avantages ?

Selon le serveur mandataire utilisé ainsi que sa configuration, les options disponibles peuvent varier, mais un proxy est capable de :

- bloquer les connexions non désirées depuis l'extérieur du réseau
- camoufler l'adresse IP et tout ce que celle-ci dévoile
- gérer les cookies envoyés par les serveurs Web
- autoriser l'accès à un serveur dont la connexion est refusée par un autre proxy
- bloquer les accès vers l'extérieur non désirés pour les membres du réseau interne
- rendre inutilisable le suivi de l'internaute (pages visitées et heures de connexion)
- accélérer la navigation grâce à la mémoire cache et le filtrage des publicités
- bloquer les publicités

Quels sont les inconvénients ?

Le serveur mandataire peut effectuer une journalisation des communications qui transitent. Puisqu'il devient le point de passage obligatoire pour toute connexion, il possède une place de choix pour la collecte d'informations sensibles comme les mots de passe, les numéros de carte de crédit, etc. Sans aller jusqu'à voler ce genre de données sensibles, un gestionnaire de proxy pourrait revendre, à des régies publicitaires, les informations collectées, rendant caduque toute tentative de protection de la vie privée. Ceci en fait une arme à double tranchant, dont il faut se méfier, et ne l'utiliser qu'en étant certain de la légitimité du service offert.

Tableau d'évaluation

Critère	Note	Justification
Installation	★★☆☆☆	Connaissance avancée nécessaire, configuration simple et rapide dans le navigateur mais recherche préalable d'un proxy nécessaire.
Répétition	★★★★★	Action définitive.
Utilisation	★★★★☆	Peut ralentir le système en fonction du proxy choisi.
Efficacité	★★★☆☆	Cache l'adresse IP et obscurcit toute l'activité.
Coût	★★★☆☆	Les proxys de qualité sont payants.
Portabilité	★★★★★	S'applique à tous les environnements.

3.5 Les VPN

Une entreprise, une organisation ou un particulier peut mettre en place son propre réseau local, dont les ressources ne sont pas partagées sur Internet, on parle dans ce cas de réseau privé. Cela ne signifie pas pour autant que les utilisateurs d'un réseau local n'ont pas accès à Internet. Il est parfois nécessaire, par exemple dans le cadre d'une entreprise autorisant le télétravail ou possédant plusieurs bâtiments, d'interconnecter deux sites distants et de mettre en place un accès depuis l'extérieur, depuis Internet, vers les données internes du réseau. Cet accès se doit d'être sécurisé et sera perçu par les machines des utilisateurs comme faisant partie du réseau local, d'où l'appellation de réseau privé virtuel ou VPN (*Virtual Private Network*) [52].

Un VPN doit donc fournir à ses utilisateurs des conditions d'exploitation et de sécurité, à travers un réseau public, identiques à celles disponibles sur un réseau privé. Il ne s'agit pas d'une technologie mais bien d'un concept, englobant habituellement les termes suivants :

> Le chiffrement : Chiffrer les données transitant sur le réseau Internet garantit qu'un tiers interceptant la communication ne pourra pas en exploiter le contenu. Un système de clés asymétriques est généralement utilisé pour la facilité d'échange des clés, le système RSA étant le plus répandu.
> L'authentification : A chaque instant, il est nécessaire de s'assurer que la machine communicante est bien celle attendue, qu'elle n'a pas été remplacée par une machine espion comme c'est le cas lors d'une attaque de type *Man-in-the-middle*[3].
> Le contrôle d'intégrité : Les messages reçus par chacun des intervenants ne doit pas avoir été modifié lors de son transit sur le réseau public. Pour cela, une clé est calculée en fonction du contenu du message transmis et est envoyée parallèlement. A la réception du message et de sa clé, le destinataire calcule la clé attendue en fonction du message qu'il a reçu. La

3. L'attaque *Man-in-the-middle* désigne, en cryptographie et sécurité informatique, une forme d'espionnage actif dans laquelle l'attaquant établit des connexions indépendantes avec les victimes et relaie des messages entre elles, les laissant croire qu'elles parlent directement l'une à l'autre à travers une connexion privée, alors qu'en réalité la conversation entière est contrôlée par l'attaquant. Ce dernier doit être capable d'intercepter tout message transitant entre les deux victimes et d'en injecter de nouveaux.

comparaison des deux clés lui indique si le message reçu est identique à celui envoyé par son correspondant ou s'il a été altéré.

> Le tunnel : Les communicants ne doivent pas se soucier de la manière dont les messages sont acheminés d'un point à l'autre. Cette abstraction est assurée par le tunnel.

De nombreuses implémentations sont possibles en fonction du niveau de sécurité requis, de la taille du réseau, des connaissances techniques des gestionnaires, etc.

Etant donné le système d'échange de clés publiques préalable à toute communication, un VPN est réservé à un groupe d'usagers déterminés par authentification. Au contraire du serveur mandataire où il est aisé d'alterner les machines communicantes, l'adresse du VPN reste généralement la même.

Un VPN peut être utilisé pour communiquer avec un serveur, c'est-à-dire qu'il peut également connecter l'internaute à un serveur mandataire. Dans ce cas, ce dernier a connaissance de la totalité des activités de l'utilisateur, si les communications qu'il est chargé de transmettre ne sont pas chiffrées, comme cela a été expliqué dans le chapitre relatif aux serveurs mandataires. L'avantage est ici que les informations transmises sont totalement invisibles pour tout intermédiaire surveillant la communication, y compris pour le fournisseur d'accès à Internet. Ceci est représenté à la figure 3.2.

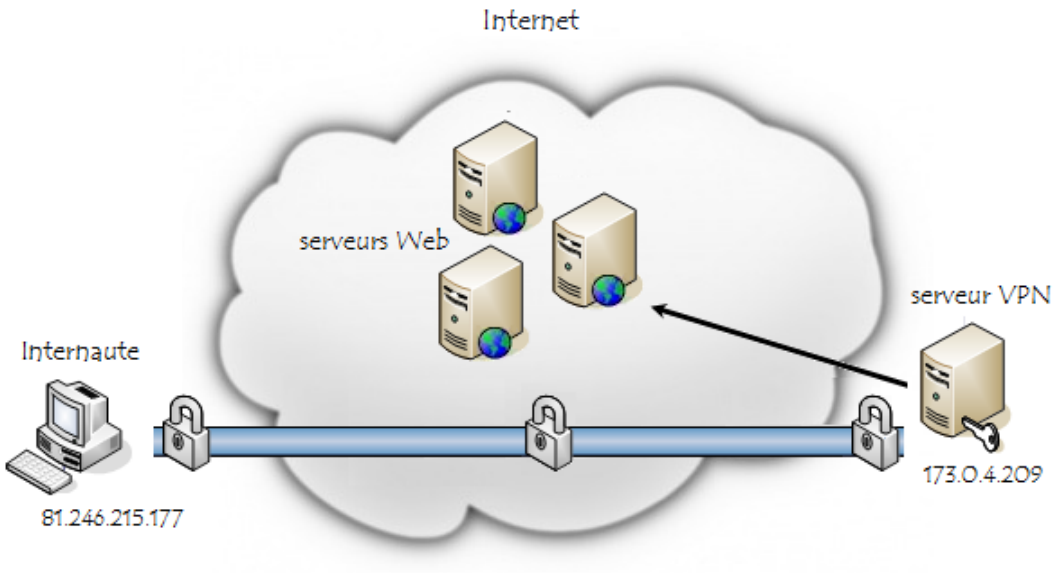


FIGURE 3.2 – Principe de fonctionnement d'un serveur privé virtuel

L'utilisation d'un VPN offre d'autres avantages par rapport à l'exploitation directe d'un serveur mandataire. Le VPN est configuré une seule fois pour la machine de l'utilisateur, tandis que le serveur mandataire doit être configuré pour chaque application séparément, certaines applications ne gérant pas cette fonctionnalité. Les programmes ne passant pas par le serveur mandataire exposeront donc directement l'adresse IP de l'internaute, et il en va de même pour les technologies Flash, Java, JavaScript et ActiveX qui offrent du contenu multimédia sur Internet mais ne prennent pas en charge l'utilisation de serveur mandataire, même si ce dernier est correctement configuré pour le navigateur Internet [53].

Un réseau privé virtuel combiné à un serveur mandataire est une solution efficace pour de

nombreux profils d'utilisateurs [54] :

- Les étudiants ou les employés d'une entreprise qui doivent pouvoir accéder aux services de l'école ou de la société, de manière sécurisée.
- Les personnes soucieuses de la protection de leur vie privée et de leur sécurité, qui considèrent que les informations qu'elles consultent ou transmettent sur Internet n'ont pas à transiter en clair et être accessibles à qui le souhaite.
- Les utilisateurs de programmes de téléchargement tels que les torrents sont fichés, cela même s'ils utilisent le service pour télécharger des fichiers légalement. L'utilisation d'un VPN leur permet de ne pas divulguer l'utilisation qu'ils font d'Internet.
- Toute personne désirant utiliser un service dont l'accès est restreint en fonction de la géolocalisation de ses clients, cette dernière ayant une préférence pour le service de proxy offert par le VPN plus que pour la sécurité qu'il confère.

Les VPN de qualité offrant un service sûr, une communication chiffrée efficacement et une communication rapide sont généralement payants. Il existe néanmoins des serveurs respectueux de la vie privée de leurs clients, dont le service est acceptable et qui permettent à l'utilisateur de se faire gratuitement sa propre opinion à propos de cette technique [55].

Quels sont les avantages ?

Le serveur virtuel privé offre la possibilité de sécuriser la connexion entre l'utilisateur et le serveur par un chiffrement des communications. De ce fait, il permet également de garantir l'identité de l'utilisateur auprès du serveur et inversement.

S'il est combiné à un serveur mandataire, il offre également tous les avantages de cette technique.

Quels sont les inconvénients ?

Les services de VPN gratuits souffrent de performances limitées, bridant de ce fait la bande passante de ses utilisateurs en formant un goulot d'étranglement. Cette limitation est inévitable au vu des coûts engendrés par une connexion de qualité et des serveurs performants.

Les services payants offrant une déclinaison gratuite, principalement à objectif de démonstration, limitent volontairement l'exploitation de leur service, afin de ne pas impacter les clients abonnés et dans l'optique de convaincre ces utilisateurs de l'intérêt à souscrire à un abonnement. Les limitations sont de ce fait liées aux performances et à la quantité de services offerts. Certaines offres brident la bande passante allouée, d'autres limitent la quantité d'information échangeable via les serveurs ou encore le temps de connexion au service. Des fonctions supplémentaires telles que la protection contre les domaines suspects ou les virus sont souvent indisponibles.

Afin de profiter d'un service de qualité et ne souffrir d'aucun ralentissement ou limitation de la quantité de données échangées sur Internet, il est donc indispensable de souscrire à un abonnement auprès d'un fournisseur de serveurs privés virtuels.

Tableau d'évaluation

Critère	Note	Justification
Installation	★★★☆☆	Beaucoup de recherches à effectuer pour trouver le VPN correspondant au but recherché, mais installation aisée.
Répétition	★★★★★	Une seule installation est nécessaire.
Utilisation	★★★★★	Peut ralentir le système en fonction du VPN choisi.
Efficacité	★★★★☆	Permet de naviguer de façon anonyme et sécurisée.
Coût	★★☆☆☆	Il faut compter un minimum de 40euros pour un VPN de qualité.
Portabilité	★★★★★	S'applique à tous les environnements.

3.6 Les réseaux informatique anonymes

Les systèmes de protection par serveur mandataire ne suffisent pas à garantir l'anonymat de leur utilisateur, les routeurs par lesquels ses requêtes transitent conservent les traces de leur passage, ce qui permet de remonter sa piste pour l'identifier. Cependant, certains systèmes sont conçus pour maintenir l'anonymat sur Internet et rendent impossible ou impraticable le suivi par adresse IP.

C'est le cas du réseau **Tor**, acronyme de "*The Onion Router*", qui désigne un réseau décentralisé et mondial de routeurs, appelés "noeuds" ou "relais", dont les communications sont cryptées et qui n'ont chacun la connaissance que du noeud précédent et du noeud suivant. Au lieu d'accéder directement au serveur désiré, la connexion passe par de nombreux relais, masquant l'adresse IP d'origine et chiffrant la communication entre l'utilisateur et le dernier relai, celui-ci a néanmoins accès aux informations mais ne sait pas qui a initié la requête.

Un observateur placé à un endroit précis de la chaîne ne pourra pas savoir d'où proviennent les données et où elles sont acheminées. De plus, la suite de noeuds à franchir est aléatoire, réduisant le risque d'analyse du trafic en répartissant les transactions en plusieurs endroits sur Internet.

Première étape : Le programme client, sur l'ordinateur de l'utilisateur, contacte un serveur d'annuaire pour obtenir une liste de noeuds Tor ainsi que leurs clés publiques.

Deuxième étape : Le programme client construit une chaîne de noeuds par lesquels transitera la requête de l'utilisateur. Celle-ci est encryptée à l'aide de la clé publique du dernier noeud Tor de la chaîne. Le résultat est ensuite encrypté avec la clé publique de l'avant-dernier noeud, après y avoir ajouté l'adresse du dernier noeud, auquel il devra transmettre la requête. Cette étape est répétée afin de construire de manière incrémentale une requête Tor couche par couche, voir figure 3.3.

Troisième étape : Le programme client transmet la requête qui a été constituée au premier noeud Tor, qui la décrypte à l'aide de sa clé privée. Il a alors accès à l'adresse IP du noeud auquel il doit transmettre la requête, et connaît naturellement l'adresse IP de la machine

qui lui a fourni la requête, ce qui lui permettra de transmettre la réponse. Il transmet donc la requête décryptée au second noeud Tor, qui la décrypte avec sa propre clé privée et en ressort les informations concernant le prochain maillon de la chaîne, ainsi de suite jusqu'au dernier noeud qui, après avoir décrypté la requête à l'aide de sa clé privée, obtient une requête non cryptée qu'il transmet au serveur cible. Il s'agit de la seule communication non cryptée de la chaîne. Le serveur répond naturellement au dernier noeud Tor, de qui semble émaner la requête. Celui-ci transmet la réponse au noeud précédent, qui fait de même, jusqu'à transmettre la réponse au programme client de l'utilisateur, qui déballe toutes les couches d'encryptage à l'aide des clés publiques, et obtient la réponse du serveur [56].

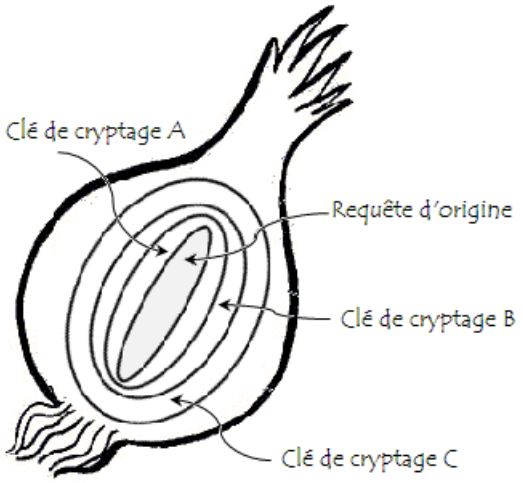


FIGURE 3.3 – Encryptage en oignon d'une requête Tor

La transmission d'une requête via le réseau Tor est illustrée à la figure 3.4.

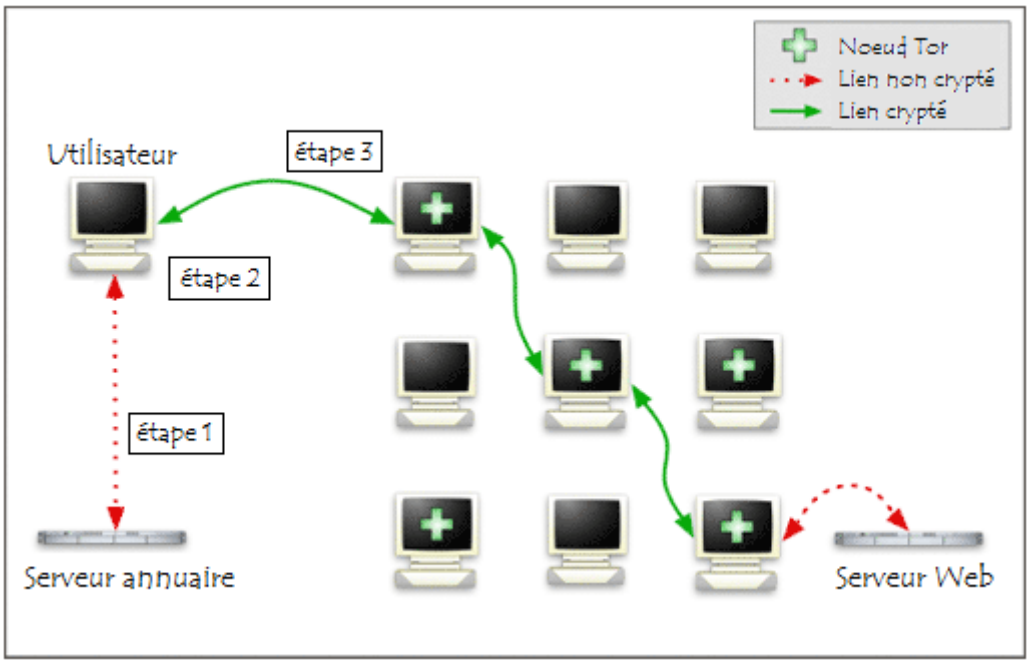


FIGURE 3.4 – Etapes d'une communication sur le réseau Tor

Tor permet ainsi de transmettre n'importe quel flux TCP (*Transmission Control Protocol*) de manière anonyme et donc de réduire sensiblement le traçage et la censure des internautes en constituant un réseau anonyme quasiment de bout en bout. La communication est, par contre, entièrement cryptée si le paquet d'origine transmis au client Tor est déjà crypté. Il s'agit alors d'HTTPS ou communication sécurisée. Les communications UDP (*User Datagram Protocol*) ne sont pas supportées par le réseau Tor et ne sont donc pas protégées, bien que ce protocole soit utilisé notamment pour les requêtes aux serveurs DNS (*Domain Name System*), qui permettent de traduire un nom de domaine en adresse IP à contacter, et sont donc nécessaires à la navigation sur Internet mais sont extrêmement révélatrices des habitudes de navigation de l'internaute.

Le réseau Tor permet d'accéder à des services cachés qui autorisent de publier des sites Internet ou d'autres services, en conservant secrète l'adresse IP et donc l'identité du serveur qui les héberge. Ces serveurs doivent s'enregistrer auprès de Tor pour recevoir une adresse d'extension ".onion" qui ne sera utilisable que par les clients du réseau Tor. Les accès au serveur suivront un protocole mis en place par Tor, qui consiste en la définition d'un rendez-vous entre le client et le serveur, sur un noeud du réseau Tor. De cette manière, le serveur et le visiteur restent tous deux anonymes.

Etant donné le fait que l'application client permettant d'exploiter le réseau Tor est intégrée au navigateur Internet, sa procédure d'installation est détaillée dans le chapitre traitant de l'installation d'extensions au navigateur.

Un autre réseau anonyme, Freenet, consiste en un disque dur géant partagé et distribué entre les ordinateurs clients. Lors de l'utilisation du réseau, celui-ci fait transiter des paquets de données chiffrés et en conserve une partie sur le disque dur de l'ordinateur, toujours sous forme chiffrée. Cette manière de procéder rend les données les plus populaires aussi les plus fréquentes, ce qui constitue un excellent moyen de résistance à la censure.

D'autres réseaux anonymes existent encore, comme I2P ou "*Invisible Internet Project*", qui fonctionne de manière analogue à Tor.

Quels sont les avantages ?

Les communications d'un internaute peuvent être chiffrées de bout en bout, lui assurant que seuls sa machine et le serveur contacté ont connaissance des données échangées.

La requête finale auprès du serveur étant effectuée par un noeud du réseau Tor, l'adresse IP de l'internaute est masquée auprès du serveur accédé, à la manière d'un serveur mandataire.

Les requêtes émises effectuent de nombreux sauts de noeud en noeud, ce qui rend le traçage des communications extrêmement compliqué, voire impossible.

Quels sont les inconvénients ?

L'utilisation de réseaux anonymes engendre une connexion beaucoup plus lente aux serveurs Internet, ceci étant dû à la recherche de noeuds et aux multiples chiffrements et déchiffrements des données.

L'anonymat est un outil à double tranchant. Les journalistes utiliseraient les services du réseau Tor pour consulter ou rédiger des articles sur la toile tout en conservant leur anonymat, mais les militaires ou organismes gouvernementaux peuvent travailler sous couverture à des fins

d'espionnage [57].

La liste des noeuds Tor est volontairement rendue publique [58], leur surveillance est donc aisée. Une attaque de type *Time Pattern*, basée sur l'observation de la fréquence de transit des paquets à travers un noeud, permet de suivre un lot de paquets d'un noeud à l'autre. Finalement, les gouvernements ayant accès aux fichiers de journalisation des fournisseurs d'accès à Internet, l'identité d'un internaute peut leur être révélée.

D'un autre côté, l'utilisation de noeuds répartis dans le monde entier rend l'identification de l'utilisateur extrêmement compliquée voire totalement impraticable. C'est la raison pour laquelle les réseaux anonymes sont parfois appelés les "réseaux d'impunité".

La liste des serveurs Tor étant connue publiquement, certains serveurs bloquent l'accès à leurs services pour les internautes passant par le réseau anonyme. C'est notamment le cas des serveurs de Wikipedia, qui interdisent l'édition d'articles si l'adresse IP du client est cachée par ce moyen.

Tableau d'évaluation

Critère	Note	Justification
Installation	★★★☆☆	Requiert l'installation d'un logiciel.
Répétition	★★★★★	Une seule configuration est nécessaire.
Utilisation	★★★☆☆	Ralentissement visible des requêtes Internet (multimédia à éviter).
Efficacité	★★★★★	Permet de naviguer de façon anonyme.
Coût	★★★★★	Gratuit.
Portabilité	★★★★★	S'applique à tous les environnements.

3.7 Les messageries électroniques

La protection des communications électroniques n'est pas chose aisée et est mise à mal dès que l'un des correspondants possède une boîte auprès d'un service non respectueux de la vie privée. Quelques solutions existent cependant pour limiter au maximum le suivi des échanges de messages électroniques.

3.7.1 Les serveurs de messagerie électronique privés

Les services de messagerie espions fournis par les géants du Web sont les plus répandus mais pas les seuls disponibles sur le marché. D'autres sociétés fournissent un service gratuit et soucieux de la vie privée, tel que GMX mail qui assure qu'il n'analysera jamais les messages de ses clients à des fins publicitaires [59]. La société Surfboard Holding BV, auteur des moteurs de recherche anonymes Ixquick et Startpage, fournira prochainement un service de messagerie privée payant nommé Startmail et qui garantira, lui aussi, la confidentialité des échanges en n'analysant aucun message et en ne journalisant que les métadonnées strictement nécessaires (voir section sur l'actualité).

Il existe également des services mis en place par des particuliers, qui n'ont à priori aucun contact avec des agences publicitaires et qui souhaitent fournir des boîtes de messagerie privées. Il faut toutefois faire preuve d'une extrême méfiance vis-à-vis de ce type de serveur, car son propriétaire dispose d'un accès total à l'ensemble des messages transitant par son service. De plus, leurs performances et disponibilités sont incomparables avec les boîtes professionnelles et sont sujettes, comme l'illustre le défunt Fritalk.com, à disparition puisque dépendantes de la volonté d'une poignée de gestionnaires.

Une dernière solution consiste à acheter un nom de domaine et souscrire à un service d'adresse électronique auprès d'un hébergeur professionnel. Ce type de service est payant mais offre à l'utilisateur la possibilité d'avoir une adresse personnalisée et un service professionnel garantissant la protection des messages, via des backups, et une disponibilité maximale du serveur. En tant que société, ils sont soumis aux lois de protection de la vie privée et ne peuvent donc pas lire personnellement les messages de leurs clients, et étant donné que leur service est payant, ils n'ont ni l'intérêt, ni bien souvent les moyens matériels, d'analyser les messages des internautes.

3.7.2 Les communications électroniques chiffrées

Les attaques par espionnage permettent à un tiers d'écouter les communications transitant sur une ligne. Les messages électroniques sont vulnérables à ces attaques, puisqu'ils circulent en clair sur le réseau. Il suffit à l'espion de se connecter sur une ligne non sécurisée, tel qu'une borne WiFi non protégée, et d'utiliser un programme d'écoute de paquets tel que WireShark pour pouvoir consulter tous les paquets non chiffrés entrants et sortants du réseau, en ce compris les messages électroniques.

Pour éviter ce désagrément, il faut faire appel à la cryptographie. Un standard de chiffrement appelé OpenPGP, est largement supporté par de nombreux clients de messagerie [60]. OpenPGP est une norme déposée par PGP Inc. en 1997 auprès de l'IETF, responsable du développement et de la promotion de standards Internet. Il existe deux suites logicielles très répandues qui implémentent le standard OpenPGP, à savoir PGP (*Pretty Good Privacy*) et GPG (*GNU Privacy Guard*). Leur utilisation n'est pas limitée au seul envoi de messages, PGP permettant de chiffrer n'importe quel fichier [61].

Les réacheminеurs de messages de type I, Cypherpunk, et de type III, Mixminion, exploitent le standard OpenPGP pour chiffrer les messages et les paquets de données. La boîte de messagerie en ligne Startmail autorisera le chiffrement des messages de ses clients selon OpenPGP. Il est également possible de chiffrer les messages transmis depuis une boîte électronique en ligne qui ne propose pas ce service dans son offre. Une extension de navigateur remplit cette fonction pour les boîtes Outlook, Gmail, Yahoo mail et GMX mail. Cette solution est présentée au chapitre relatif à l'installation d'extensions au navigateur.

Parallèlement à cela, le chiffrement permet de s'assurer que seul le destinataire légitime d'un message pourra le lire à l'aide de sa clé privée, et fournit également à ce dernier l'assurance que le message qu'il a reçu provient bien de l'expéditeur mentionné dans l'en-tête. Comme expliqué au point précédent traitant des serveurs de messagerie anonymes, il est aisé de remplacer les données contenues dans l'en-tête d'un message électronique, en ce compris l'adresse de l'expéditeur, et ainsi se faire passer pour une tierce personne.

Le service de messagerie privée GMX mail fournit un chiffrement à l'aide de SSL (*Secure Socket Layer*) mais ce service n'est pas comparable à PGP. S'ils travaillent tous deux avec un système de clés, PGP travaillant avec un système hybride clé symétrique - clé asymétrique, SSL

travaillant avec un système de clé symétrique, ils n'ont pas la même vocation. Là où PGP excelle dans le chiffrement de fichiers et de textes, SSL a pour objectif la sécurisation d'une communication, en travaillant avec des certificats, contenant notamment la clé de chiffrement mais également d'autres informations, et garantissant l'identité des deux communicants. Il est d'ailleurs envisageable d'établir une communication sécurisée à l'aide de SSL, en HTTPS (*HyperText Transfer Protocol Secure*), et d'y faire transiter des messages chiffrés à l'aide d'OpenPGP [62].

3.7.3 Les serveurs de messagerie anonymes

Les solutions d'anonymisation concernant les messages électroniques exposées jusqu'à présent n'apportent qu'une protection relative de l'identité car, si elles sont efficaces face à la grande majorité des services publicitaires et des particuliers, elles le sont moins pour des personnes ayant une connaissance des technologies de communication, qui pourront recouper les adresses IP utilisées pour envoyer différents messages, et elles sont totalement inefficaces vis-à-vis des autorités qui pourront, en partenariat avec le fournisseur d'accès à Internet, remonter très facilement à l'identité de l'auteur des messages. Seuls les serveurs de messagerie anonymes fournissent une véritable protection de la vie privée en permettant de cacher tant le contenu des messages que l'identité de leur expéditeur et leur destinataire.

Un *remailer*, réacheminеur de messages ou encore serveur de messagerie anonyme est un serveur ayant pour rôle la réception de courriers et leur expédition vers le destinataire final. Le message reçu doit pour cela contenir des instructions notamment concernant les destinataires du message à transmettre. Le service retire au passage de l'en-tête du message toute information permettant de remonter à l'expéditeur, il ne préserve que l'objet du message. Seul l'opérateur du service peut retrouver les autres traces et si une chaîne de plusieurs remailers est établie, il devient difficile de retrouver ces informations. Leur utilisation est comparable à celle d'un serveur mandataire (proxy), qui réachemine les paquets de données.

L'en-tête des messages est une information généralement peu utile à l'utilisateur et lui est donc cachée par la plupart des clients de messagerie. Cet en-tête contient en effet les informations relatives au transit du message telles que l'adresse de messagerie de l'expéditeur et les noeuds Internet que le message a traversé. Le message est composé de paquets de données, contenant chacun l'adresse IP des noeuds expéditeurs et destinataires du paquet, ces informations ne pouvant pas être supprimées. Les réacheminеurs de messages masquent donc évidemment l'adresse de messagerie de l'expéditeur mais également son adresse IP ainsi que les adresses de chaque noeud traversé par les paquets de données. Cette liste d'adresses IP est remplacée par une fausse, tandis que l'adresse IP d'origine du message est remplacée par celle du serveur de réacheminement.

Il existe plusieurs types de serveurs de messagerie anonymes [63] :

- Les **pseudonymous** remplacent l'adresse de messagerie de l'expéditeur du message par un pseudonyme, puis transmettent le message au destinataire. Ce dernier pourra ainsi répondre au message, qui sera réceptionné par le réacheminеur et transmis à l'utilisateur du service.

- Les types I ou **Cypherpunk** réceptionnent des messages chiffrés (d'où l'appellation cypherpunk) contenant l'adresse du destinataire. Ils déchiffrent ensuite le message pour récupérer l'adresse du destinataire, remplacent l'adresse de l'expéditeur et transmettent le message. Un espion ne pourra donc pas connaître le contenu du message échangé. Il ne sera pas non plus possible au destinataire de répondre au message. Une chaîne de réacheminеurs de messages de type I est comparable au réseau Tor, où chaque noeud déchiffre le message

pour découvrir l'adresse du noeud suivant, sans connaître l'expéditeur du message, son destinataire ni son contenu.

- Les types II ou **Mixmaster** transmettent les messages sous forme de paquets de taille fixe dont l'ordre est permuté (d'où l'appellation Mixmaster), empêchant ainsi tout espionnage des messages entrants et sortants du réacheminеur. Le type Mixmaster peut être combiné avec le type Cypherpunk, pour une sécurité accrue.

- Les types III ou **Mixminion** découpent les messages en paquets de taille fixe. Pour chaque paquet, un chemin est défini, composé de serveurs de mix. Chaque paquet est ensuite chiffré à l'aide des clés publiques de chaque serveur qui sera traversé. Les paquets sont ensuite transmis séparément vers le premier serveur mix de leur chemin, qui les déchiffre via sa clé privée pour découvrir l'adresse du serveur mix suivant dans la chaîne ou le destinataire final du message si la chaîne est terminée. Chaque serveur de mix ne connaît que ses serveurs adjacents dans la chaîne et ignore tout de l'expéditeur, du destinataire et du message transité. Le principe de fonctionnement des réacheminеurs de messages de type Mixminion est similaire en de nombreux points à celui du réseau Tor. De plus, tout comme ce dernier, Mixminion nécessite l'installation d'un programme client sur la machine de l'utilisateur.

Certains sites Internet proposent d'envoyer anonymement des courriers électroniques, jouant le rôle de pseudonymous et rendant de ce fait l'exploitation de leur service aussi aisée que celle d'une boîte de messagerie classique. Nonobstant l'absence de quelques facilités, telles que la liste des contacts ou l'historique des messages envoyés, plusieurs de ces sites combinent les serveurs de messagerie électronique anonymes et les boîtes de messagerie jetables. Ils permettent dès lors à leurs clients, lorsqu'ils complètent le formulaire d'envoi d'un message électronique, d'indiquer une adresse de messagerie jetable ainsi qu'une véritable adresse, qui sera invisible pour le destinataire du message, mais vers laquelle la réponse de ce dernier sera transmise [64]. Un exemple de site Internet offrant une boîte de messagerie de type pseudonymous est présent à l'annexe C.

Une illustration de l'envoi d'un message à un réacheminеur de type II est disponible à l'annexe D.

3.7.4 Techniques diverses

Des techniques plus originales peuvent être envisagées, comme par exemple la transformation du texte à envoyer en une image, à l'aide d'une capture d'écran. Toute analyse du texte du message est dès lors rendue impraticable. Une telle technique n'est cependant pas automatisable comme les techniques de chiffrement et ne protège pas les autres informations telles que les identités de l'expéditeur et du destinataire, le sujet du message, les pièces jointes, ou encore l'en-tête.

Un algorithme de chiffrement basique tel le chiffrement dit "de César" pourrait être convenu entre les deux communicants. Il s'agit alors de remplacer les lettres de l'alphabet par d'autres, toujours les mêmes. Néanmoins cette technique requerrait un énorme effort cognitif à ses utilisateurs et consisterait à appliquer un chiffrement moins efficace que l'OpenPGP.

Quels sont les avantages ?

Les messageries électroniques privées offrent la garantie que le contenu, les pièces jointes et le titre des messages n'est pas surveillé, aucune personne et aucun algorithme n'analyse les messages du client. Par contre, les contacts et la fréquence de communication avec chacun d'entre eux seront bientôt surveillés par obligation légale (voir section sur l'actualité).

Les messageries électroniques privées fournies par des particuliers garantissent l'absence d'algorithme d'analyse des messages.

Les boîtes de messagerie associées à un nom de domaine garantissent un service professionnel exempt d'analyse des communications, la confidentialité, la protection, la sauvegarde régulière des messages ainsi que la disponibilité du service sont assurés. Il est de plus possible d'obtenir plusieurs boîtes de messagerie pour un seul nom de domaine, ce qui en limite les coûts.

Le service fourni par les réacheminеurs de messages est le seul à offrir un anonymat complet, en ce compris l'adresse IP. Les autorités consultant la journalisation du service auront la possibilité de connaître les heures d'envoi des messages mais pas leur sujet, leur contenu ni leurs destinataires, s'il s'agit d'une chaîne de réacheminеurs.

Le chiffrement des courriers électroniques assure qu'un espion surveillant la ligne n'aura aucune connaissance du contenu du message, garantissant à l'expéditeur que seul le destinataire pourra lire le message. De plus, le destinataire a l'assurance que le message reçu provient effectivement de l'expéditeur mentionné dans l'en-tête du message.

Quels sont les avantages ?

Les messageries électroniques privées fournies par des particuliers sont très risquées, le ou les gestionnaires ont accès à toutes les informations et les contenus des messages, il n'y a aucune garantie de sécurité ou de performance et il demeure toujours un risque de disparition soudaine du service.

Les boîtes de messagerie associées à un nom de domaine sont toujours payantes. Il faut enregistrer un nom de domaine et souvent souscrire à un service d'hébergement pour obtenir une ou plusieurs boîtes de messagerie associées au nom de domaine.

Les réacheminеurs de messages sont complexes et nécessitent un effort constant de la part de leur utilisateur.

Le chiffrement des courriers électroniques nécessite que les correspondants possèdent chacun un client de messagerie gérant le PGP.

Finalement, un inconvénient général à l'utilisation de boîtes de messagerie électronique est que l'internaute peut fournir autant d'effort qu'il le souhaite, il sera toujours tributaire de la fiabilité des boîtes de messagerie de ses correspondants. Il suffit qu'un seul destinataire d'un message exploite un service qui analyse et journalise le contenu des messages pour que l'ensemble des correspondants de la communication soient impactés. Se protéger de l'espionnage à ce niveau est donc très compliqué.

Tableaux d'évaluation

Les serveurs de messagerie électronique privés :

Critère	Note	Justification
Installation	★★★★☆	Nécessite l'inscription à une nouvelle boîte de messagerie ou l'abonnement à un service d'hébergement.
Répétition	★★★★★	Action définitive.
Utilisation	★★☆☆☆	Contraintes dûes au changement de boîte de messagerie / dépendance possible de la qualité de service d'un particulier.
Efficacité	★★★☆☆	Plus d'analyse des messages / surveillance possible par un particulier.
Coût	★★★★☆	Gratuit ou bon marché.
Portabilité	★★★★★	S'applique à tous les environnements.

Les serveurs de messagerie anonymes :

Critère	Note	Justification
Installation	★★★★★	Rien ne doit être installé.
Répétition		Non applicable.
Utilisation	★★☆☆☆	Action à réitérer à chaque envoi de message.
Efficacité	★★★★☆	Message chiffré et adresse IP masquée pour la meilleure solution.
Coût	★★★★★	Gratuit.
Portabilité	★★★★★	S'applique à tous les environnements.

3.8 Le rejet des applications mobiles trop intrusives

Afin d'éviter d'être tracé lors de chaque déplacement, téléphone en poche, il est recommandé de ne pas activer les options de géolocalisation proposées sur les *smartphones* et les tablettes. Ces options sont présentées comme des améliorations du comportement de certaines applications, sans en préciser les détails, mise à part pour les applications de navigation, où il est indiqué que l'activation de la carte réseau sans fil permet d'accélérer la localisation de l'utilisateur, par rapport à la seule utilisation de la puce GPS.

Néanmoins, lorsqu'on y prête attention, de nombreuses applications demandent le droit d'accès à la position géographique de l'utilisateur sans en préciser la raison. Si l'utilisateur a activé les options de géolocalisation sur son appareil, ces informations sont envoyées de manière invisible

pour celui-ci, toute application peut donc potentiellement accéder aux coordonnées de l'utilisateur, en temps réel.

Un autre problème spécifique aux appareils mobiles est le téléchargement et l'installation d'applications requérant un accès illégitime aux ressources du téléphone. La liste des droits d'accès requis par une application est affichée à l'écran avant de procéder à son installation. Il est de la responsabilité de l'utilisateur de parcourir cette liste et de décider en connaissance de cause, s'il accepte d'autoriser l'application à recevoir les accès qu'elle demande.

Quels sont les avantages ?

L'utilisateur peut gérer la transmission de ses informations de localisation. Lorsqu'une application tente d'accéder à ces données, et si l'utilisateur n'a pas activé les services de localisation sur son appareil, il en sera notifié et pourra décider s'il autorise ou non l'application à prendre connaissance de sa position géographique.

L'utilisateur gère les applications installées sur son appareil et donc indirectement les informations qu'il autorise à divulguer.

Quels sont les inconvénients ?

Les applications demandant l'accès aux informations de localisation refuseront de démarrer leur service si l'accès à cette information leur est interdit.

Tableau d'évaluation

Critère	**Note**	**Justification**
Installation	★★★★★	Rien ne doit être installé ou configuré.
Répétition	★☆☆☆☆	Décision à prendre pour chaque application et à chaque exécution concernant la géolocalisation.
Utilisation	★★☆☆☆	Certaines applications sont refusées totalement.
Efficacité	★★★★☆	Les informations personnelles et la localisation sont masquées.
Coût	★★★★★	Gratuit.
Portabilité	★☆☆☆☆	Ne s'applique qu'aux *smartphones*.

3.9 La configuration du navigateur

De nombreuses améliorations peuvent être apportées aux paramètres par défaut du navigateur Internet afin de limiter les traces laissées sur la toile mais également sur l'ordinateur de l'internaute. En effet, de nombreuses informations sont enregistrées sur son ordinateur, le plus souvent à son insu : cookies, cookies flash, historique de navigation, formulaires, mots de passe, fichiers téléchargés, fichiers temporaires, etc.

La configuration du navigateur peut dès lors automatiser le nettoyage de ces informations et également empêcher une partie de l'espionnage lors de la navigation.

L'installation d'extensions permet de grandement améliorer les options de filtre et de nettoyage proposées par les navigateurs. Ces extensions étant développées par des particuliers ou des organismes indépendants des navigateurs, il en existe plusieurs centaines et leur nombre est en constante augmentation, c'est pourquoi seules les plus réputées, les plus efficaces et les plus originales d'entre elles seront abordées [65].

Il est finalement recommandé de varier l'utilisation des navigateurs, en réservant spécifiquement chacun d'entre eux pour une tâche définie, ce qui ne dispense pas d'appliquer les recommandations de configuration et d'extensions :

Internet Explorer : navigateur incontournable et indétrônable, il reste le plus utilisé de nos jours et est installé systématiquement sur chaque ordinateur utilisant le système d'exploitation Windows [66]. C'est la raison pour laquelle une grande partie des sites gouvernementaux sont uniquement garantis compatibles pour ce dernier. Il est donc judicieux de réserver ce navigateur pour ce type de services, principalement l'utilisation de la carte d'identité électronique ou eID.

Firefox : navigateur développé par une communauté et non par une société et de ce fait plus respectueux de la vie privée de ses utilisateurs. Disposant d'une grande communauté active, de nombreuses extensions sont disponibles et il est sujet à de fréquentes mises-à-jour. Il est recommandé d'exploiter ce navigateur pour la navigation quotidienne et les recherches sur Internet. Le navigateur Opera est également développé indépendamment de toute société mais représente seulement 1,5% des parts de marché et dispose donc, logiquement, d'une communauté moins nombreuse et d'une quantité plus restreinte d'extensions.

Chrome : autre excellent navigateur, dont l'utilisation est statistiquement juste inférieure à Firefox, disposant de nombreuses possibilités de personnalisation. Cependant, étant donné son appartenance à l'entreprise Google qui a déjà prouvé par le passé qu'elle n'hésite pas à empiéter sur la vie privée de ses clients (voir section sur l'actualité), il est recommandé de limiter au maximum son utilisation. L'entreprise américaine, très active sur Internet, fournit de nombreux services de qualité qui sont généralement parfaitement intégrés à son navigateur. Ce denier peut donc être exploité dans le cadre de l'utilisation de services nécessitant une connexion de l'utilisateur, ces services ne se limitant pas à ceux fournis par Google mais également par tout service dont l'entreprise collecte des informations relatives à l'activité de ses clients.

Cette répartition des activités entre plusieurs navigateurs permet de garantir à l'internaute que le navigateur qu'il utilise pour accéder à des services gouvernementaux est exempt de logiciels espions et compatible avec ces services, que le navigateur choisi pour sa navigation quotidienne ne pistera pas ses déplacements et ses moindres faits et gestes, et finalement il pourra obtenir ces avantages sans se priver de l'utilisation des services de son choix, y compris s'ils sont fournis par de grandes entreprises qui collectent ses données.

3.9.1 Configuration de base

Tout navigateur enregistre des traces de l'activité Internet de ses utilisateurs. Certaines d'entre elles autorisent des fonctionnalités, tels que les cookies, tandis que d'autres ont pour vocation l'amélioration de l'expérience de l'utilisateur, ce qui est le cas des fichiers temporaires. Ces traces ne sont pas supprimées automatiquement à la fin de la session de navigation et peuvent révéler de nombreuses informations relatives à l'utilisateur, ce qui pourrait l'amener à désirer supprimer ces traces, tout particulièrement s'il a utilisé un ordinateur public.

Parmi ces informations, bon nombre peuvent être supprimées en une simple opération ou tout simplement automatiquement à la fermeture du navigateur. La **navigation privée** est proposée par la plupart des navigateurs et a pour vocation, non pas de naviguer de manière anonyme, mais bien d'être assuré de ne laisser aucune trace sur l'ordinateur. La même opération peut être effectuée manuellement ou encore automatiquement lorsque le navigateur est fermé. Toutes les informations ne sont pas supprimées, mais les principales sont effectivement nettoyées, comprenant :

- ◇ L'historique de navigation : Il s'agit de la liste des sites visités par l'utilisateur, ainsi que les dates et heures de dernière consultation.
- ◇ Les formulaires : Chaque champ texte complété pour effectuer une recherche ou s'identifier auprès d'un service est enregistré par le navigateur, pour pouvoir lui être proposé à nouveau lors de sa prochaine visite.
- ◇ Les mots de passe : Le navigateur peut être configuré pour ne jamais enregistrer les mots de passe, toujours demander l'avis de l'utilisateur ou toujours enregistrer les mots de passe. Dans ce dernier cas, aucun avertissement ne lui sera présenté.
- ◇ Les cookies : Présentés dans un chapitre précédent, ces petits fichiers permettent d'identifier l'internaute entre deux sessions de navigation. Leur consultation permet d'avoir connaissance des domaines visités par ce dernier.
- ◇ Les favoris : Il s'agit de raccourcis menant vers les sites que l'utilisateur a indiqué comme marque-pages.

Les cookies flash nécessitent une opération particulière car, n'étant pas encore gérés par l'ensemble des navigateurs, il sera parfois nécessaire de les supprimer manuellement. Ils sont enregistrés dans les dossiers "adobe" et "macromedia" du dossier personnel de l'utilisateur.

La révélation de certaines informations relatives à la navigation auprès de divers serveurs peut être évitée en configurant quelques options du navigateur :

- L'option ***Do Not Track*** ou **ne pas me pister** est un standard mis en place par l'organisme W3C et qui permet à l'internaute d'indiquer aux sites visités ses préférences concernant l'analyse de son activité [67]. Cette option, désormais proposée par la majorité des navigateurs, y compris Internet Explorer, Firefox, Safari et Chrome, reste toutefois symbolique, les sites restent libres de respecter ou non notre requête.
- La configuration de la politique de gestion des cookies permet d'accepter ou de refuser ces derniers, de choisir leur durée de stockage via leur date d'expiration, de consulter les cookies actifs sur une page Internet en saisissant "javascript :alert(document.cookie)" dans la barre d'adresse du navigateur, et finalement de spécifier les sites ayant l'autorisation de créer des cookies, étant donné que le rejet complet des cookies rend de nombreux sites inutilisables, tels que les paniers d'achat de magasins en ligne ou les sites exigeant une connexion à l'aide d'identifiants.
- Le refus systématique des cookies tierce partie.
- La désactivation de la géolocalisation, disponible sous la référence "geo.enabled" des op-

tions avancées, en entrant "about :config" dans la barre de navigation de Firefox ou "about :flags" pour Chrome. Cette option ne permet évidemment pas d'empêcher toute localisation, l'adresse IP restant toujours présente, mais elle indique au navigateur qu'il n'a pas l'autorisation de transmette les informations de localisation qu'il pourrait détenir, comme le nom du réseau sans fil auquel l'ordinateur est connecté.
- Désactiver l'envoi des pages consultées précédemment, faisant partie des informations contenues dans les en-têtes HTTP, en forçant la référence "network.hhtp.sendRefererHeader" des options avancées à la valeur 0, en entrant "about :config" dans la barre de navigation de Firefox ou "about :flags" pour Chrome.
- Désactiver l'exécution de la technologie Flash, utilisée pour créer des cookies LSO et pour afficher des publicités interactives.
- Désactiver le JavaScript, qui permet d'accéder aux informations de l'ordinateur et du navigateur et de les transmettre au serveur fournissant le script.

Finalement, le navigateur Firefox présente un avantage non négligeable en intégrant un outil à sa barre d'adresse, servant de "moteur de recherche" interne à l'historique de navigation. Cela lui permet de consulter les pages enregistrées sur le disque dur de l'utilisateur, via l'historique de navigation et, si une correspondance avec les critères de recherche est trouvée, d'afficher la page concernée sans établir aucune communication avec un quelconque serveur distant.

Afin de répondre aux craintes légitimes des internautes, le projet P3P (*Platform for Privacy Preference*) a été démarré par l'organisme W3C pour permettre aux sites de communiquer de manière standardisée leur charte de confidentialité. Cette dernière, étant présente dans l'en-tête HTTP de la réponse du serveur, elle peut être automatiquement récupérée et interprétée par le navigateur, ce qui lui permettra de la communiquer à l'utilisateur et de prendre automatiquement les décisions adéquates en fonction des préférences de ce dernier. Cette solution permet à l'internaute de naviguer sans devoir lire la charte de confidentialité de chaque site Internet [68].

Quels sont les avantages ?

Toute trace laissée sur l'ordinateur peut être supprimée automatiquement ou manuellement sur décision de l'utilisateur. La suppression de ces données peut se faire de manière sélective, il est par exemple possible de supprimer l'historique de navigation en conservant les cookies, ou encore de supprimer uniquement les mots de passe enregistrés.

Nombre d'informations transmises involontairement lors de la navigation peuvent être conservées anonymes en indiquant simplement au navigateur que les sites ne disposent pas de l'autorisation d'y accéder.

Quels sont les inconvénients ?

La désactivation de certaines options nuit aux performances de services spécifiques, par exemple le refus de transmettre les informations de géolocalisation ne permet plus aux sites de localiser l'internaute et donc de personnaliser leur service. D'autres options peuvent empêcher tout accès à un site ou une fonctionnalité, comme le rejet des cookies ou encore la désactivation de la technologie Flash.

En effaçant systématiquement ses traces, l'utilisateur se prive d'une partie du confort offert par les navigateurs. Il ne pourra plus retrouver un site visité lors d'une session de navigation précédente, il devra également retenir ses identifiants pour chaque site.

Tableau d'évaluation

Critère	Note	Justification
Installation	★★★☆☆	Les options ne sont pas concentrées dans un endroit unique, nécessité d'effectuer des recherches.
Répétition	★★★★★	Action unique.
Utilisation	★★★★☆	Diminution du confort d'utilisation.
Efficacité	★★☆☆☆	Quantité limitée d'informations cachées lors de la navigation.
Coût	★★★★★	Gratuit.
Portabilité	★★★★★	S'applique à tous les environnements.

3.9.2 Configuration avancée via l'installation d'extensions

La protection de la vie privée est un sujet d'actualité dont de plus en plus d'internautes se soucient. Aussi, si les navigateurs Internet commencent à fournir quelques options de configuration, les extensions de navigateurs foisonnent et certaines peuvent réellement avoir un impact positif sur la qualité des sessions de navigation des internautes et sur la sécurité de leurs données personnelles. Dans ce chapitre, seront abordées les extensions les plus célèbres et les plus utiles, mais il faut toutefois garder à l'esprit qu'il en existe souvent plusieurs réalisant le même travail et que certaines d'entre elles se chevauchent inévitablement. Il ne s'agit donc bien évidemment pas d'une liste exhaustive mais d'une approche informative des capacités de tels outils.

L'installation d'extensions requiert une attention particulière de la part de l'utilisateur car il n'y a pas de contrôle de la qualité et de la légitimité des extensions proposées sur le site du navigateur. Il est donc possible qu'une extension se fasse passer pour un outil de protection de la vie privée ou toute autre application utile pour l'internaute, mais d'espionner les habitudes de navigation de l'utilisateur ou encore ses données privées, et de les transmettre à un serveur Internet.

Certaines extensions sont susceptibles d'empêcher l'utilisation correcte d'un site Internet, en bloquant ses services totalement ou partiellement. La majorité des extensions fournissent une option de désactivation par site, ou peuvent simplement être désactivées temporairement.

FIGURE 3.5 – Icônes des navigateurs les plus célèbres. De gauche à droite : Mozilla Firefox, Internet Explorer, Google Chrome, Apple Safari et Opera

Cette analyse se concentrera sur le navigateur Firefox assurant privilégier la protection de la vie privée de ses utilisateurs. Ses extensions [69] sont souvent compatibles avec d'autres navigateurs et, lorsque ce n'est pas le cas, ces derniers possèdent souvent leurs propres extensions réalisant les mêmes fonctions. Leur énumération ne ferait qu'altérer la lisibilité et la clarté de l'analyse. Néanmoins, les navigateurs supportés par les extensions présentées seront désignés à l'aide de leur logo, dont la liste est reprise à la figure 3.5.

Adblock plus 

Extension très polyvalente qui permet à l'internaute de décider du comportement des sites Internet qu'il visite. Il travaille à partir de dizaines d'abonnements à des listes noires régulièrement mises à jour auxquelles l'utilisateur peut souscrire et qui peuvent être complétées par ses soins en un simple clic de souris sur un élément indésirable affiché lors de sa navigation. Adblock peut également être configuré individuellement pour chaque site visité. L'extension fournit les options suivantes :

- Bloqueur de publicités offrant la possibilité à l'internaute de bloquer les domaines indésirables tels que les sites publicitaires. Par défaut, cette extension ne bloque que les publicités intrusives, par exemple générant une fenêtre intempestive (pop-up) mais l'utilisateur pourra la configurer pour bloquer l'entièreté des publicités. Cette option permet à l'utilisateur de supporter les sites dont la survie dépend de la rémunération offerte par l'affichage des publicités.
- Bloqueur de virus et logiciels espions en interdisant à l'utilisateur l'accès aux domaines qui sont réputés pour diffuser des logiciels malveillants.
- Suppression des boutons des medias sociaux, supprimant des sites visités les boutons des réseaux sociaux qui espionnent les habitudes de navigation.
- Bloqueur de pixels espions mis en place par des sociétés publicitaires, permettant d'éviter leur affichage sur les pages visitées et de ce fait l'installation des cookies tierce partie sur l'ordinateur de l'internaute.
- Bloqueur de scripts pouvant empêcher l'exécution de scripts Java et Flash.

HTTPS Everywhere 

Extension permettant au navigateur de se connecter aux serveurs via une connexion sécurisée en utilisant le protocole SSL. De nombreux sites Internet supportent le chiffrement mais rendent son usage compliqué. Par exemple, ils utilisent par défaut une communication non chiffrée par HTTP, ou encore ils placent sur leurs pages sécurisées des liens vers la partie non sécurisée du site. HTTPS-Everywhere permet à l'utilisateur d'avoir l'assurance d'utiliser, dès que cela est possible, une connexion chiffrée pour communiquer avec le serveur.

WOT (Web of Trust) 

Extension n'ayant aucun impact sur la navigation. Elle affiche une icône colorée indiquant si le site Internet visité est globalement considéré comme digne de confiance ou non. Cette catégorisation est effectuée par les internautes eux-mêmes, qui partagent leurs expériences de navigation pour enrichir la base de données de WOT. Cette manière de fonctionner a pour avantage de laisser toujours le contrôle à l'internaute et de l'avertir à propos de menaces que les logiciels de protection ne peuvent détecter, tels que de faux sites de vente en ligne ou des attaques par hameçonnage consistant à imiter le plus fidèlement possible l'interface d'un site auquel l'internaute fait entièrement confiance et divulgue des informations privées.

User Agent Switcher 

Extension donnant le droit à l'internaute de spécifier la valeur de l'information contenue dans les en-têtes HTTP transmis aux serveurs Internet et de choisir quel navigateur Internet il semblera utiliser. Cette astuce lui permettra de passer inaperçu en choisissant d'afficher un navigateur très répandu tel qu'Internet Explorer ou encore de faire varier son empreinte unique.

Disconnect 

Extension regroupant plusieurs outils qui permettent de couper les communications entre le navigateur et les serveurs d'espionnage de Google, Facebook, Twitter, Yahoo !, Linkedin mais également les moins célèbres. Ces communications concernent les publicités, l'espionnage des habitudes de navigation et les boutons des réseaux sociaux, elles ralentissent le chargement des pages, collectent des données concernant l'internaute pour les revendre ensuite à des sociétés publicitaires. Disconnect offre également des options d'affichage de compteurs de requêtes bloquées pour chaque domaine et permet d'afficher un graphique représentant les connexions entre le site Internet visité et les domaines espions.

Request Policy 

Extension qui bloque toute requête inter-site effectuée lors de la navigation. Ce type de requête survient lorsque le site Internet visité par l'internaute génère lui-même une requête auprès d'autres domaines. Ces requêtes résultent la plupart du temps en l'affichage de publicités, de pixels espions ou encore l'exécution de scripts d'espionnage. L'extension exploite une politique bloquante qui indique à l'utilisateur que des requêtes inter-sites ont été stoppées, lui permet d'en consulter la liste et de choisir lesquelles sont légitimes et donc autorisées.

TrackMeNot 

Extension permettant d'assombrir les recherches effectuées par l'internaute sur le moteur de recherche de son choix. Travaillant de manière invisible pour l'utilisateur, elle génère des recherches aléatoires sur les moteurs de recherche AOL, Yahoo !, Google et Bing lors de chaque recherche. Il est donc impossible de tracer les recherches légitimes de l'utilisateur pour le profiler. Cette technique a pour inconvénient de surcharger chaque accès à un moteur de recherche, ce qui peut ralentir les connexions lentes et risque, si cette technique se répand, de surcharger également les moteurs de recherche.

Remove Google Tracking 

Extension spécifique aux recherches effectuées sur le moteur de recherche Google, comme son nom l'indique. Lors de l'accès à un lien affiché dans la liste de résultats, l'internaute est dirigé vers un serveur de Google, qui enregistre l'information spécifiant que ce dernier a sélectionné ce résultat spécifique, après avoir effectué une recherche comportant tels mots clés. L'internaute est ensuite redirigé du serveur de Google vers le serveur hébergeant le site qu'il désire consulter. Remove Google Tracking permet à l'internaute d'accéder directement au site visité, tel qu'indiqué dans les résultats de recherche, sans passer préalablement par les serveurs de Google. Une autre extension, nommée Remove Yahoo Tracking fonctionne de manière similaire et est destinée à stopper le suivi des recherches effectuées sur Yahoo.

Facebook Auto-Logout 

Extension offrant la possibilité à l'internaute d'être automatiquement déconnecté des services de Facebook lorsqu'il ferme son navigateur Internet ou lorsqu'il a quitté le site de Facebook depuis un laps de temps défini et configurable. Pour cela, l'extension supprime tous les cookies générés par Facebook de l'ordinateur de l'internaute, ce qui rend l'espionnage de sa navigation impossible pour le réseau social. Cette astuce permet également de ne plus se soucier de la déconnexion du service, comme expliqué au chapitre concernant la déconnexion systématique.

Mailvelope 

Extension du navigateur interagissant avec les boîtes de messagerie électroniques en ligne telles que Gmail, Outlook, Yahoo ! mail ou encore GMX mail, elle propose à ses utilisateurs de chiffrer leurs courriers électroniques à l'aide d'Open PGP et travaille donc avec un système de clé publique et clé privée.

La première étape consiste donc à générer une paire de clés en entrant un mot de passe, qui consistera en la clé privée de l'utilisateur, la clé publique est générée automatiquement par l'extension, affichée à l'utilisateur et la paire est enregistrée par l'extension du navigateur. L'internaute pourra également ajouter une clé publique pour chacun de ses contacts utilisant le chiffrement par Open PGP. L'échange de clé publique peut s'effectuer de manière libre, la clé publique n'étant, par définition, pas une donnée sensible.

Lorsqu'il rédige un nouveau message, l'internaute peut apercevoir une icône représentant un cadenas, lui permettant de signaler qu'il désire chiffrer son message. L'application utilise alors la clé publique liée au destinataire du message, qui aura été ajoutée au préalable à la base de connaissances de l'extension, chiffre le message et remplace le texte lisible par le message chiffré. Lorsque l'internaute reçoit un message chiffré, Mailvelope le détecte et affiche à nouveau une icône représentant un cadenas. Lorsque l'internaute clique dessus, il est invité à entrer sa clé privée, c'est-à-dire le mot de passe qu'il avait entré lors de la génération des clés. L'extension remplace alors le message chiffré par le texte lisible, dans l'interface de la boîte de messagerie, comme si le message avait été envoyé en clair.

TorButton

Extension qui permet d'exploiter le réseau anonyme Tor de manière transparente. Si cette extension était auparavant disponible dans le catalogue d'extensions de Firefox, ce n'est à présent plus le cas, étant donné la fréquence des mises à jour des versions du navigateur et les moyens limités de la communauté Tor, il a été décidé qu'il faudra désormais utiliser une version spécifique de Firefox pour pouvoir utiliser l'extension. La communauté Tor fournit un ensemble préconfiguré nommé Tor Browser Bundle [70] et contenant une version portable du navigateur, c'est-à-dire ne nécessitant aucune installation, seulement une extraction de l'archive téléchargée, contenant bien évidemment l'extension TorButton mais également HTTPS-Everywhere et NoScript. L'utilisateur peut donc directement démarrer une session de navigation anonyme de bout en bout, grâce à l'utilisation du réseau anonyme Tor et à une communication chiffrée en HTTPS. S'il le désire, l'internaute peut également ajouter d'autres extensions de son choix, comme Adblock plus. Par défaut, le navigateur portable est également configuré pour utiliser le moteur de recherche anonyme Start page et il fonctionne exclusivement en mode de navigation privée, c'est-à-dire qu'aucune trace de la navigation n'est enregistrée sur l'ordinateur de l'internaute.

Il existe des centaines d'extensions disponibles pour les navigateurs Internet et parmi celles-ci, des dizaines ont pour vocation la protection de la vie privée. Les solutions présentées précédemment couvrent une grande partie de la protection qu'il est possible de mettre en place, mais d'autres astuces existent probablement et de nouvelles extensions encore plus poussées seront certainement développées. L'abondance d'extensions disponibles permet également à l'utilisateur d'effectuer ses propres choix en fonction des possibilités de chacune et de la manière dont elles fonctionnent et interagissent avec celui-ci. Parmi ces autres extensions performantes mais redondantes avec celles présentées dans ce document, se trouvent Ghostery, Better Privacy, PrivacyFix, NoScript, FlashBlock, etc.

Quels sont les avantages ?

Les extensions permettent à l'internaute de personnaliser son expérience de navigation à l'extrême. Des possibilités aussi nombreuses que variées ont fait l'objet de développements, il suffit généralement à l'utilisateur de trouver les extensions qui satisfont à ses attentes.

Grâce aux extensions, chaque utilisateur est en mesure de protéger sa vie privée de manière efficace, quel que soit son niveau de connaissance de l'informatique et des techniques de profilage. En effet, les programmes sont généralement préconfigurés et leur utilisation est simplifiée au maximum.

Finalement, ces outils permettent d'automatiser des tâches qui demanderaient une attention constante de la part des internautes, et même un travail régulier vis-à-vis de la surveillance de l'activité des sites qu'il visite.

Quels sont les inconvénients ?

Certaines extensions travaillent sur le principe des listes blanches et ont un comportement bloquant par défaut. Il est parfois nécessaire d'enquêter pour identifier celle qui est responsable d'un problème d'affichage d'un site.

Une politique d'apprentissage peut requérir l'attention régulière de l'utilisateur pour connaître sa préférence pour la situation rencontrée. Cela peut s'avérer très intrusif lorsque l'extension interroge l'utilisateur plusieurs fois par chargement de page, voire carrément ingérable si plus d'un programme applique cette politique.

L'internaute doit rester méfiant quant aux extensions qu'il installe, et doit s'assurer que le service effectue strictement ce qu'il annonce. Pour cela, il peut se baser sur l'appréciation des autres utilisateurs et sur leur caractère open source du programme.

Tableau d'évaluation

Critère	Note	Justification
Installation	★★☆☆☆	Travail de recherche, de comparaison et éventuellement de configuration.
Répétition	★★★★★	Action unique.
Utilisation	★★★☆☆	Certaines extensions peuvent perturber le fonctionnement de sites ou requérir l'attention de l'utilisateur pour une action particulière.
Efficacité	★★★★☆	Nombreuses options (supprimer les redirections, cookies et publicités - sécuriser les communications)
Coût	★★★★★	Gratuit.
Portabilité	★★★★★	S'applique à tous les environnements.

3.10 Le contournement du *Yield Management*

Les achats effectués en ligne sont parfois soumis à la technique du *Yield Management*, comme cela a été expliqué dans les chapitres précédents. Voilà pourquoi il est conseillé de réaliser les simulations et les achats sur des terminaux séparés, principalement lorsqu'il s'agit de réservations de services, plus sujettes à l'application de cette technique de marketing que l'achat de biens.

Le site Internet de l'opérateur ne doit pas être en mesure de reconnaître le visiteur. Ce dernier peut donc appliquer les techniques exposées tout au long de ce chapitre pour rendre sa navigation invisible ou pour nettoyer toute trace de son passage. Il doit pour cela effacer les cookies et cookies LSO, changer son adresse IP en redémarrant son modem ou camoufler celle-ci via un serveur mandataire ou un réseau privé virtuel et finalement éviter toute exposition de caractéristiques spécifiques via les en-têtes HTTP, car certaines de ces informations peuvent influencer les prix. Par exemple, le prix d'une réservation peut augmenter en fonction du système d'exploitation embarqué sur l'ordinateur du visiteur [71].

3.11 Récapitulatif des techniques de défense

Afin d'offrir un aperçu général des informations personnelles protégées par chaque technique abordée au long de ce chapitre, un tableau comparatif a été établi, sur base des mêmes catégories que celles définies en fin du chapitre relatif aux données révélées, et qui sont, rappelons-le, fonction des données auxquelles elles donnent accès :

- **Identité réelle** : Les informations légales relatives à l'internaute, permettant de l'identifier, telles que définies dans la loi "vie privée".
- **Identité fictive** : Toute information permettant de reconnaître l'internaute entre deux sessions de navigation, et d'établir un profil le concernant, sans pour autant connaître son identité réelle.
- **Données sensibles** : Toute information strictement confidentielle pour l'internaute, tels que les mots de passe et les numéros de carte de crédit.
- **Données personnelles** : Informations relatives à l'identité réelle de l'internaute mais ne permettant pas de déduire cette identité ni d'effectuer une reconnaissance entre deux connexions (âge, sexe, langue, etc).
- **Localisation** : La position géographique approximative de l'internaute, voire son adresse postale.
- **Centres d'intérêt** : Les thèmes qui tiennent à coeur à l'internaute et pourront être exploités principalement dans le cadre de publicités ciblées.
- **Pages visitées** : La liste complète ou une partie des sites Internet visités avec l'adresse, la date, l'heure, etc.
- **Matériel utilisé** : Reconnaissance de l'environnement de connexion de l'internaute, comme son ordinateur, la version du système d'exploitation, le type de navigateur, etc.

	Identité réelle	Identité fictive	Données sensibles	Données personnelles	Localisation	Centres d'intérêt	Pages visitées	Matériel utilisé
La déconnection systématique							■	
Les moteurs de recherche anonymes						■	■	
Les identités virtuelles	■							
Les proxys	■				■			
Les VPN	■		■		■			
Les réseaux informatiques anonymes	■		■		■			
Les messageries électroniques	■			■		■		
Le rejet des applications mobiles trop intrusives				■	■			
La configuration du navigateur	■	■	■	■		■	■	■

Chapitre 4

Analyse des techniques de respect de la vie privée à usage des gestionnaires

De nos jours, il n'existe pas une organisation qui n'enregistre d'information relative à ses clients. Les gestionnaires de ces sites Internet ou plus largement les responsables de tout service en ligne sont tenus de protéger au mieux et de manière adéquate les données des utilisateurs, dont ils ont la charge. Le côté juridique fut abordé à la section traitant des devoirs de gestionnaires, dans le chapitre de mise en contexte. Abordons maintenant brièvement quelques bonnes pratiques qui peuvent aider les gestionnaires à maintenir une protection efficace des données qui leur sont confiées. Toute politique de sécurité nécessite une analyse préalable des risques, de leur probabilité et de l'impact de leur survenance, mais aussi de la nature des informations à protéger [72].

4.1 Les données

Certaines bonnes pratiques sont également des obligations légales vis-à-vis de la loi vie privée (voir chapitre de mise en contexte). Il est donc obligatoire de ne récolter que les informations strictement nécessaires à l'exécution du traitement et également d'obscurcir les données des clients après expiration du traitement prévu initialement mais aussi en cas de traitement réalisé par une personne.

Le gestionnaire veillera également à ne jamais sauvegarder de manière non chiffrée des informations qui ne doivent pas être consultables. Il est par exemple plus que recommandé de ne pas enregistrer les mots de passe en clair, car ceux-ci doivent pouvoir être comparés à un équivalent chiffré, mais pas consultés.

Pour éviter tout désagrément tel que celui illustré par le fait d'actualité relatif à la divulgation accidentelle d'informations par la SNCB, il est judicieux de ne placer aucune donnée privée quelle qu'elle soit sur un serveur accessible depuis Internet, si ces informations ne sont pas susceptibles de devoir être accédées elles-mêmes par Internet.

4.2 Les protocoles de communications

Des protocoles de communication sécurisés existent et ont été éprouvés. C'est le cas par exemple du protocole SSL permettant d'effectuer une connexion sécurisée de type HTTPS entre deux communicants, par un système d'échange de clés.

Si le type de communication s'y prête, un tunnel VPN peut être mis en place entre le serveur et le client, afin d'assurer la sécurité de toute information transitant par Internet.

4.3 Les cookies

Déjà présentés précédemment, les cookies sont de petits fichiers stockés sur les ordinateurs des visiteurs et rendant possible l'enregistrement des informations permettant de les identifier d'une visite à l'autre. La structure des cookies et les informations qu'ils contiennent sont entièrement laissées à l'appréciation des gestionnaires de site. De nombreuses failles de sécurité peuvent provenir de ces petits fichiers, il incombe donc au responsable du site d'assurer la sécurité de ses clients. Voici quelques bonnes pratiques à prendre en compte lors de la définition de la politique de gestion des cookies [73] :

- Il ne peut pas contenir d'information d'authentification, y compris de manière chiffrée.
- Enregistrer un ID de session, généré de manière aléatoire en début de session.
- Configurer les cookies de manière à ce qu'ils expirent en fin de session et soient donc supprimés automatiquement par le navigateur du visiteur.
- Protéger son système contre les ń replay attack ż qui consistent à introduire manuellement un cookie, côté client, contenant un ID de session ayant déjà servi, de manière à rejouer la session. Pour éviter ce type d'attaque, il faut invalider les ID distribués aux visiteurs, en fin de session.
- Chiffrer le contenu des cookies, particulièrement s'ils contiennent une information compréhensible par un humain ou une machine autre que le serveur du site.
- Ne jamais utiliser le mécanisme des cookies pour reconnaître automatiquement un visiteur, sans nécessiter d'authentification. Un tiers ayant récupéré le cookie d'un client pourrait alors usurper son identité auprès du site.

Conclusion

La protection de la vie privée est un sujet brûlant. Les articles traitant du sujet et exposant problèmes et solutions diverses prolifèrent dans les journaux en ligne. Cette prise de conscience collective et l'intérêt croissant qui en découle pour la préservation des données confidentielles permettent à de nombreuses solutions techniques de voir le jour. Les utilisateurs se voient alors proposer des outils de sécurité fournis par des sociétés ayant identifié une opportunité commerciale. Des organisations à but non lucratif proposent également des produits gratuits développés par une communauté de passionnés. Ceux-ci permettent souvent aux internautes d'apporter leur pierre à l'édifice, par exemple en participant à la diffusion du produit, comme c'est le cas pour les noeuds Tor, ou à son amélioration, comme c'est le cas des développeurs d'extensions pour les navigateurs.

Il est à la portée de chacun d'appliquer diverses techniques afin de protéger sa vie privée. Les premières étapes à mettre en place ne consistent d'ailleurs pas en l'installation d'outils mais elles nécessitent une modification des habitudes de l'internaute. Celui-ci doit adopter certains réflexes, comme changer la page d'accueil de son navigateur, se déconnecter des services après les avoir utilisés, se méfier des programmes qu'il télécharge et installe en s'assurant qu'il ne s'agit pas de logiciels espions, et finalement éviter au possible de centraliser toutes ses activités auprès d'une même société, en choisissant des solutions alternatives. Il est de bon ton d'exploiter des services respectueux de la vie privée, comme que le moteur de recherche DuckDuckGo ou la boîte de messagerie GMX mail. Une astuce permet à l'internaute de maîtriser son image sur Internet en se créant une ou plusieurs identités virtuelles. Finalement, le réglage de quelques paramètres du navigateur Internet permet, combiné à l'ajout de l'adresse d'un serveur mandataire, de terminer cette liste de techniques de défense des données qui, remarquons le, ne nécessitent aucune installation de la part de l'internaute et également aucune contrainte supplémentaire dans l'utilisation des services Internet.

S'il désire aller plus loin dans le niveau de protection, l'utilisateur doit alors mettre la main à la pâte et installer un navigateur alternatif à ceux fournis par les grandes firmes, comme Mozilla Firefox et y ajouter des extensions permettant d'améliorer encore sa protection en empêchant par exemple l'exécution de scripts et l'enregistrement de cookies.

La grande variété des outils disponibles sur Internet rend possible, quelques soient les compétences et les moyens financiers de la personne, la mise en place rapide d'une protection efficace. Chaque outil a un impact limité sur la quantité et la qualité des informations que l'internaute divulgue lors de ses sessions de navigation mais, une fois combinés, ils permettent de protéger la moindre trace que l'utilisateur laisse dans son sillage.

Finalement, pour atteindre une protection maximale des données, y compris auprès des autorités et du fournisseur d'accès à Internet, l'internaute dispose de solutions de chiffrement et d'anonymisation de ses communications, au moyen des VPN, des réseaux informatiques anonymes tels que Tor, et des serveurs de messagerie anonymes. La protection des informations à

caractère personnel est un droit qu'il est encore possible de faire valoir si l'on s'en donne les moyens. Aucune solution globale ne peut, pour le moment, être mise en place par les autorités, ceci dû au manque de standards, au retard des règlementations au sujet du monde du numérique, ainsi qu'à la nécessité de faire collaborer des gouvernements du monde entier. L'attente passive de ces solutions n'est donc pas envisageable, il est de la responsabilité de chacun d'agir pour défendre les informations qu'il juge importantes.

L'anonymat sur Internet, au sens propre, est impossible à atteindre. Des informations nécessaires au fonctionnement de toute communication doivent transiter sur le réseau et laissent de ce fait des traces du passage de tout individu. Ces traces peuvent toutefois être tellement complexes à remonter qu'il sera en pratique impossible d'identifier leur émetteur. Cet anonymat, bien que relatif, est néanmoins très risqué car il peut être exploité à des fins criminelles. Il est primordial, selon moi, de pouvoir identifier une personne sur Internet comme dans tout autre lieu public, où chaque citoyen est tenu de présenter sa carte d'identité à un agent de la fonction publique. D'un autre côté, les Etats-Unis, sous prétexte de surveillance des activités terroristes, depuis le 11 septembre 2001, ne cessent de tenter de contrôler la moindre activité des citoyens du monde entier. Il est essentiel de trouver un compromis entre surveillance et vie privée, et ne pas tomber dans l'excès auquel nous assistons de nos jours.

Bibliographie

[1] Ball J., "NSA's Prism surveillance program : how it works and what it can do", The Guardian, 8 juin 2013. Consulté le 03/08/2013 sur http://www.theguardian.com/world/2013/jun/08/nsa-prism-server-collection-facebook-google

[2] Lee T.B., "Here's everything we've learned about how the NSA's secret programs work", Washington Post, 25 juin 2013.
Consulté le 03/08/2013 sur http://www.washingtonpost.com/blogs/wonkblog/wp/2013/06/25/heres-everything-weve-learned-about-how-the-nsas-secret-programs-work/

[3] Greenwald G., "XKeyscore : NSA tool collects "nearly everything a user does on the internet"",
The Guardian, 31 juillet 2013. Consulté le 03/08/2013 sur http://www.theguardian.com/world/2013/jul/31/nsa-top-secret-program-online-data

[4] Gellman B. and Poitras L., "U.S., British intelligence mining data from nine U.S. Internet companies in broad secret program", Washington Post, 6 juin 2013.
Consulté le 03/08/2013 sur http://articles.washingtonpost.com/2013-06-06/news/39784046_1_prism-nsa-u-s-servers

[5] Stroobants J.-P., "Prism aurait permis d'empêcher un attentat en Belgique"
sur http://www.lemonde.fr/europe/article/2013/06/20/prism-aurait-permis-d-empecher-un-attentat-en-belgique_3433149_3214.html, consulté le 03/08/2013.

[6] RTBF, "Tous vos e-mails bientôt stockés pendant un an : et la vie privée ?" sur http://www.rtbf.be/info/belgique/detail_tous-vos-emails-bientot-stockes-pendant-un-an?id=8043908, consulté le 07/08/2013.

[7] Le Soir, "Les échanges de mails stockés un an" sur http://www.lesoir.be/276922/article/actualite/belgique/2013-07-08/echanges-mails-stockes-un-an, consulté le 07/08/2013.

[8] CPVP, "Questions les plus fréquemment posées - Google Street View" sur http://www.privacycommission.be/fr/faq-page/388#t388n7448, consulté le 07/08/2013.

[9] Google Maps, "Street View : confidentialité" sur http://maps.google.be/intl/fr/help/maps/streetview/privacy.html, consulté le 03/08/2013.

[10] Kravets D., "An Intentional Mistake : The Anatomy of Google's Wi-Fi Sniffing Debacle" sur http://www.wired.com/threatlevel/2012/05/google-wifi-fcc-investigation, consulté le 03/07/2013.

[11] Oeillet A., "Street View : Google condamné en Allemagne pour violation de la vie privée" sur http://pro.clubic.com/entreprises/google/actualite-555190-google-condamne-allemagne-violation-vie-privee.html, consulté le 03/08/2013.

[12] L'avenir, "Fuite de donnée à la SNCB : la Commission vie privée informe les clients concernés" sur http://www.lavenir.net/article/detail.aspx?articleid=DMF20130328_00289092, consulté le 12/07/2013.

[13] CPVP, "La Commission vie privée informe à propos des développements dans l'affaire de la fuite de données à la SNCB" sur http://www.privacycommission.be/en/node/8453, consulté le 12/07/2013.

[14] RTBF, "Fuite de données à la Défense : une "erreur technique" rectifiée vendredi" sur http://www.rtbf.be/info/belgique/detail_fuite-de-donnees-a-la-defense-une-erreur-technique-rectifiee-vendredi?id=7901546, consulté le 12/07/2013.

[15] CNIL, "Vos traces > Les moteurs de recherche" sur http://www.cnil.fr/vos-droits/vos-traces/les-moteurs-de-recherche/, consulté le 20/08/2013.

[16] de Vries L., "CIA Caught Sneaking Cookies" sur http://www.cbsnews.com/2100-205_162-504131.html, consulté le 16/07/2013.

[17] Wikipedia, "Daniel Brandt" sur http://fr.wikipedia.org/wiki/Daniel_Brandt, consulté le 15/08/2013.

[18] Poullet Y. & Rouvroy A., "Le droit à l'autodétermination informationnelle et la valeur du développement personnel : une réévaluation de l'importance de la vie privée pour la démocratie", Etat de droit et virtualité, Montréal, Thémis, 2009. - pp. 157-222.

[19] Kessous E., Mellet K. & Zouinar M., "L'économie de l'attention : entre protection des ressources cognitives et extraction de la valeur", Sociologie du Travail, vol. 52, n° 3, 2010, p.359-373.

[20] Susan B. Barnes, "A privacy paradox : Social networking in the United States", journal "First Monday" du 4 septembre 2006.

[21] Le Monde, "Internet : un décret impose aux hébergeurs de conserver les mots de passe" sur http://www.lemonde.fr/technologies/article/2011/03/02/internet-un-decret-impose-aux-hebergeurs-de-conserver-les-mots-de-passe_1487396_651865.html?xtmc=internet&xtcr=160, consulté le 21/06/2013.

[22] Cert-IST, "Obligations des entreprises pour la journalisation des connexions" sur http://www.cert-ist.com/fra/ressources/Publications_ArticlesBulletins/Autres/obligation_legal_journalisation/, consulté le 01/07/2013.

[23] Freyssinet E., "Décret d'application de la LCEN sur la conservation des données par les FAI et hébergeurs" sur http://blog.crimenumerique.fr/2011/03/04/decret-dapplication-de-la-lcen-sur-la-conservation-des-donnees-par-les-fai-et-hebergeurs/, consulté le 10/08/2013.

[24] Leila Brahim, "Obligation de conservation des données pour les entreprises" sur http://www.e-juristes.org/Obligation-de-conservation-des/, consulté le 07/06/2013.

[25] CNIL, "Vos traces > Votre ordinateur" sur http://www.cnil.fr/vos-droits/vos-traces/votre-ordinateur, consulté le 21/07/2013.

[26] RIPE Network Coordination Centre, site officiel http://www.ripe.net/, consulté le 02/07/2013.

[27] CNIL, "Vos traces > Les cookies" sur http://www.cnil.fr/vos-droits/vos-traces/votre-ordinateur, consulté le 21/07/2013.

[28] Wikipedia, "Cookie (informatique)" sur http://fr.wikipedia.org/wiki/Cookie_%28informatique%29, consulté le 02/07/2013.

[29] Wikipedia, "Local shared object" sur http://en.wikipedia.org/wiki/Local_shared_object, consulté le 02/07/2013.

[30] D. Robinson & K. Coar, "The Common Gateway Interface (CGI) Version 1.1" sur http://www.ietf.org/rfc/rfc3875, consulté le 19/08/2013.

[31] The Apache Software Foundation, "Apache et les variables d'environnement" sur https://httpd.apache.org/docs/trunk/fr/env.html, consulté le 19/08/2013.

[32] Microsoft Developer Network, "IIS Server Variables" sur http://msdn.microsoft.com/en-us/library/ms524602(v=vs.90).aspx, consulté le 19/08/2013.

[33] CNIL, "Vos traces > L'historique" sur http://www.cnil.fr/vos-droits/vos-traces/votre-ordinateur, consulté le 21/07/2013.

[34] CNIL, "Vos traces > Les moteurs de recherche" sur http://www.cnil.fr/vos-droits/vos-traces/votre-ordinateur, consulté le 21/07/2013.

[35] The Radicati Group, Inc., "Email Statistics Report, 2010-2014" sur http://www.radicati.com/?p=5282, consulté le 20/08/2013.

[36] Julien L., "Google accusé de scanner le contenu des emails" sur http://www.numerama.com/magazine/17411-google-accuse-de-scanner-le-contenu-des-emails.html, consulté le 02/08/2013.

[37] Google, "Règles de confidentialité du service Gmail" sur https://mail.google.com/mail/help/about_privacy.html, consulté le 01/07/2013.

[38] Fleishman G., "How the iPhone knows where you are" sur http://www.macworld.com/article/1159528/how_iphone_location_works.html, consulté le 10/08/2013.

[39] Barnier M., Soriano P., Gratadour J.-R. & Plat O., "l'e-commerce transfrontière - l'europe numérique au coeur des échanges" sur http://www.associationeconomienumerique.fr/wp-content/uploads/2012/01/Synth%C3%A8se-livre-ecommerce-transfronti%C3%A8re.pdf, consulté le 19/08/2013.

[40] M. Bourdin J., "Commerce électronique : l'irrésistible expansion", Rapport d'information n° 272 (2011-2012) http://www.senat.fr/rap/r11-272/r11-272_mono.html, consulté le 19/08/2013.

[41] CNIL, "Cloud computing" sur http://www.cnil.fr/les-themes/technologies/cloud-computing/browse/1/, consulté le 10/07/2013.

[42] Vion-Dury P., "Facebook, Gmail : le "cloud computing" met-il en danger notre vie privée ?" sur http://leplus.nouvelobs.com/contribution/226562-facebook-gmail-le-cloud-computing-met-il-en-danger-notre-vie-privee.html, Le nouvel Observateur, consulté le 10/07/2013.

[43] Zebuzzeo, "Évitez le Cloud Computing Si Vous Tenez à Votre Vie Privée" sur http://zebuzzeo.blogspot.be/2013/07/evitez-le-cloud-computing-si-vous-tenez.html, consulté le 27/07/2013.

[44] Microsoft, "Why do I need to add security info ?" sur http://windows.microsoft.com/en-US/windows-live/account-security-password-information, consulté le 20/08/2013.

[45] Korben, "Google Chrome et son spyware" sur http://korben.info/google-chrome-et-son-spyware.html, consulté le 21/06/2013.

[46] Lamandé E., "PETs : pour une anonymisation et souveraineté des données personnelles..." sur http://www.globalsecuritymag.fr/PETs-pour-une-anonymisation-et,20090619,10381.html, consulté le 20/08/2013.

[47] DuckDuckGo, "Ne laissez plus de traces !" sur http://donttrack.us/, consulté le 21/06/2013.

[48] Lucie Ronfaut, "Google plus puissant que jamais dans la recherche" sur http://www.lefigaro.fr/hightech/2013/07/09/01007-20130709ARTFIG00326-google-plus-puissant-que-jamais-dans-la-recherche.php, consulté le 14/07/2013.

[49] Coquis C., "DuckDuckGo et Startpage : les moteurs de recherche anti-Google" sur http://articles.softonic.fr/duckduckgo-et-startpage-les-moteurs-de-recherche-anti-google, consulté le 21/06/2013.

[50] site de http://proxy.org/, consulté le 21/08/2013.

[51] Communauté Wikibooks, "Comment contourner un proxy" sur http://fr.wikibooks.org/wiki/Comment_contourner_un_proxy, consulté le 18/06/2013.

[52] De Reynal D., De Rorthais J.G & Tan S.S., "Présentation sur les VPN", Université de Marne-la-Vallée, http://monge.univ-mlv.fr/~duris/NTREZO/20032004/DeReynal-DeRorthais-Tan-VPN.pdf, Février 2004

[53] Greg, "La différence entre un proxy et un VPN" sur http://desgeeksetdeslettres.com/hardware/difference-proxy-vpn-anonyme#ixzz2bxpJ6EqI, consulté le 8/07/2013.

[54] Henry A., "Why You Should Start Using a VPN (and How to Choose the Best One for Your Needs" sur http://lifehacker.com/5940565/why-you-should-start-using-a-vpn-and-how-to-choose-the-best-one-for-your-needs, consulté le 8/07/2013.

[55] Enigmax & Ernesto, "VPN Services That Take Your Anonymity Seriously, 2013 Edition" sur http://torrentfreak.com/vpn-services-that-take-your-anonymity-seriously-2013-edition-130302/, consulté le 8/07/2013.

[56] WikiLeaks, "Tor" sur http://www.wikileaks.org/wiki/WikiLeaks:Tor, consulté le 10/06/2013.

[57] Tor Project, "Who Uses Tor" sur https://www.torproject.org/about/torusers.html.en, consulté le 10/06/2013.

[58] Tor Project, "Tor Network Status" sur http://torstatus.blutmagie.de/, consulté le 10/06/2013.

[59] GMX, "Keep your account clean" sur http://www.gmx.com/, consulté le 11/07/2013.

[60] OpenPGP Alliance, "About OpenPGP" sur http://www.openpgp.org/about_openpgp, consulté le 9/08/2013.

[61] Wikipedia, "Pretty Good Privacy" sur http://en.wikipedia.org/wiki/Pretty_Good_Privacy, consulté le 9/08/2013.

[62] Nicolas, "SSL/PGP, Modèles de confiance" sur http://www.kns7.org/informatique/4/13/SSL/PGP,-Modeles-de-Confiance.html, consulté le 9/08/2013.

[63] Wikipedia, "Anonymous remailer" sur http://en.wikipedia.org/wiki/Anonymous_remailer, consulté le 10/07/2013.

[64] Arobase.org, "L'e-mail anonyme" sur http://www.arobase.org/securite/email-anonyme.htm, consulté le 10/07/2013.

[65] Tor Project, "What is the Tor Browser Bundle ?", sur https://www.torproject.org/projects/torbrowser.html.en#windows, consulté le 12/07/2013.

[66] NetMarketShare, "Desktop Browser Market Share" sur http://www.netmarketshare.com/browser-market-share.aspx?qprid=0&qpcustomd=0, consulté le 12/07/2013.

[67] Bonvoisin G., "Do Not Track : le standard anti-pistage du W3C" sur http://www.cnetfrance.fr/news/do-not-track-le-standard-anti-pistage-du-w3c-39765631.htm, consulté le 13/07/2013

[68] Platform for Privacy Preferences Project, "Enabling smarter Privacy Tools for the Web" sur http://www.w3.org/P3P/, consulté le 13/07/2013

[69] Mozilla Foundations, "Firefox Add-ons" sur https://addons.mozilla.org/fr/firefox/, consulté le 12/07/2013.

[70] Henry A., "The Best Browser Extensions that Protect Your Privacy" sur http://lifehacker.com/the-best-browser-extensions-that-protect-your-privacy-479408034, consulté le 12/07/2013.

[71] Amsili S., "Un site propose des hôtels plus chers aux utilisateurs de Mac" sur http://www.lefigaro.fr/conso/2012/06/26/05007-20120626ARTFIG00480-un-site-propose-des-hotels-plus-chers-aux-utilisateurs-de-mac.php?cmtpage=0, consulté le 29/07/2013

[72] CPVP, "Une politique de sécurité de l'information" sur http://www.privacycommission.be/fr/politique-securite-information, consulté le 27/08/2013

[73] Jacques A., "Cookies et sécurité" sur http://www.securiteinfo.com/conseils/cookies.shtml, consulté le 27/08/2013

Annexes

A Exemples d'exploitation de l'adresse IP

De nombreux sites permettent d'avoir connaissance de son adresse IP et d'être grossièrement localisé sur une carte, la précision étant d'environ 40km. Par exemple, en effectuant une recherche sur le site "http ://whatismyipaddress.com/fr/mon-ip" via un ordinateur situé à Vilvorde, celui-ci renseigne la région de Bruxelles comme illustré à la figure A.1.

FIGURE A.1 – WHOIS effectué sur base d'une adresse IP

Certains navigateurs Internet permettent une localisation beaucoup plus précise, à quelques mètres près dans certaines zones fortement peuplées.

Par exemple, Firefox peut rassembler les informations sur les points d'accès sans fil alentour en plus de l'adresse IP de l'ordinateur et les envoyer au fournisseur de service de géolocalisation par défaut, "Google Location Services", pour faire une estimation de localisation de l'utilisateur. Cette estimation de localisation peut alors être partagée avec des sites Web avec l'accord de ce dernier (https ://www.mozilla.org/fr/firefox/geolocation/).

Une nouvelle recherche, réalisée du même endroit, mais avec les paramètres de géolocalisation activés, fournit des résultats beaucoup plus précis comme on peut le constater à la figure A.2.

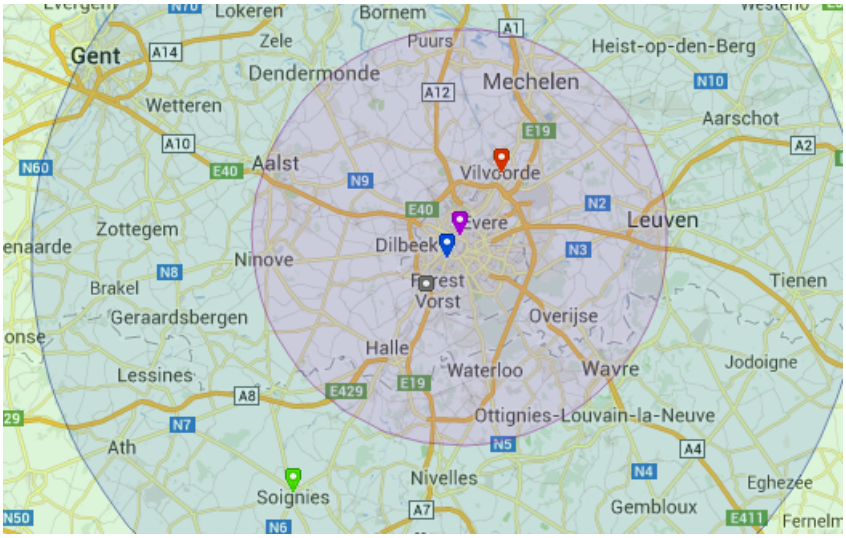
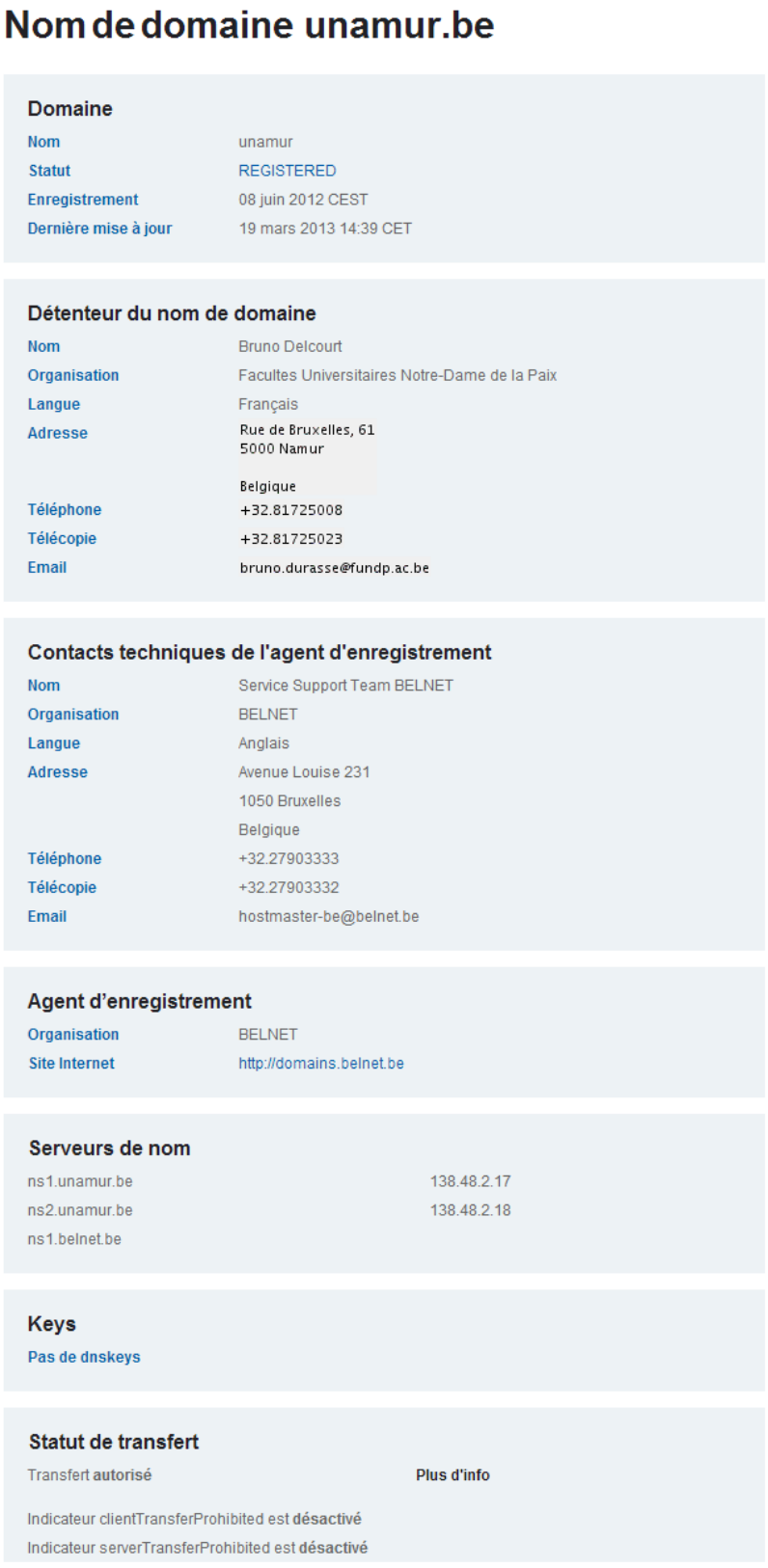


FIGURE A.2 – Géolocalisation effectuée sur base des données du navigateur Internet et de différents services de géolocalisation (W3C Geolocation, Quova, Maxmind et IP2Location)

Le service WHOIS peut renseigner nombre d'informations sur un nom de domaine comme illustré à la figure A.3 avec le nom de domaine de l'Université de Namur.



Nom de domaine unamur.be

Domaine

Nom	unamur
Statut	REGISTERED
Enregistrement	08 juin 2012 CEST
Dernière mise à jour	19 mars 2013 14:39 CET

Détenteur du nom de domaine

Nom	Bruno Delcourt
Organisation	Facultes Universitaires Notre-Dame de la Paix
Langue	Français
Adresse	Rue de Bruxelles, 61 5000 Namur Belgique
Téléphone	+32.81725008
Télécopie	+32.81725023
Email	bruno.durasse@fundp.ac.be

Contacts techniques de l'agent d'enregistrement

Nom	Service Support Team BELNET
Organisation	BELNET
Langue	Anglais
Adresse	Avenue Louise 231 1050 Bruxelles Belgique
Téléphone	+32.27903333
Télécopie	+32.27903332
Email	hostmaster-be@belnet.be

Agent d'enregistrement

Organisation	BELNET
Site Internet	http://domains.belnet.be

Serveurs de nom

ns1.unamur.be	138.48.2.17
ns2.unamur.be	138.48.2.18
ns1.belnet.be	

Keys

Pas de dnskeys

Statut de transfert

Transfert **autorisé** **Plus d'info**

Indicateur clientTransferProhibited est **désactivé**

Indicateur serverTransferProhibited est **désactivé**

FIGURE A.3 – WHOIS effectué sur base du nom de domaine www.unamur.be

Enfin, ce service peut également être utilisé sur un *smartphone* et retourne les résultats affichés à la figure A.4.

FIGURE A.4 – WHOIS effectué sur base de l'adresse IP d'un terminal mobile

B Exemple d'échange de cookies avec le serveur Google

De nombreux outils permettent d'analyser les transactions HTTP qui s'effectuent entre un navigateur et un serveur Web. On utilise ici le module de Firefox *Live HTTP Headers*.

Requête HTTP envoyée lors d'une première connexion, c'est-à-dire sans cookies, au serveur Google Belgique à l'aide de l'URL "www.google.be" :

```
GET / HTTP/1.1
Host : www.google.be
User-Agent : Mozilla/5.0 (Windows NT 6.1 ; WOW64 ; rv :22.0) Gecko/20100101 Firefox/22.0
Accept : text/html,application/xhtml+xml,application/xml ;q=0.9,*/* ;q=0.8
Accept-Language : fr,fr-fr ;q=0.8,en-us ;q=0.5,en ;q=0.3
Accept-Encoding : gzip, deflate
Connection : keep-alive
```

Et la réponse de Google :

```
HTTP/1.1 200 OK
Cache-Control : private, max-age=0
Content-Encoding : gzip
Content-Type : text/html ; charset=UTF-8
Date : Fri, 26 Jul 2013 16 :34 :35 GMT
Expires : -1
P3P : CP="This is not a P3P policy !
See http ://www.google.com/support/accounts/bin/answer.py ?hl=enanswer=151657 for more info."
Server : gws
Set-Cookie : PREF=ID=1f148233d9ea654f :FF=0 :TM=1374856475 :LM=1374856475 :S=_oSFdzEgenk-
FXNm ; expires=Sun, 26-Jul-2015 16 :34 :35 GMT ; path=/ ; domain=.google.be
Set-Cookie : NID=67=hqYOMaBqiwg8VmmrnJ5uEIWXhxtMxwNhVYiX2NcErsB2O26rMhIuToQFG
5nUPxwXYLoVBkNVInFinBae54YvWbMlfmewN7m_3elajzilWKmns6EnQZbl3 HZnJOiHmzs0 ;
expires=Sat, 25-Jan-2014 16 :34 :35 GMT ; path=/ ; domain=.google.be ; HttpOnly
X-Frame-Options : SAMEORIGIN
X-XSS-Protection : 1 ; mode=block
X-Firefox-Spdy : 3
```

Seconde connexion à www.google.be, avec cookie cette fois-ci :

```
GET / HTTP/1.1
Host : www.google.be
User-Agent : Mozilla/5.0 (Windows NT 6.1 ; WOW64 ; rv :22.0) Gecko/20100101 Firefox/22.0
Accept : text/html,application/xhtml+xml,application/xml ;q=0.9,*/* ;q=0.8
Accept-Language : fr,fr-fr ;q=0.8,en-us ;q=0.5,en ;q=0.3
Accept-Encoding : gzip, deflate
Cookie : PREF=ID=1f148233d9ea654f :FF=0 :TM=1374856475 :LM=1374856475 :S=_oSFdzEgenk-
FXNm ; NID=67=hqYOMaBqiwg8VmmrnJ5uEIWXhxtMxwNhVYiX2NcErsB2O26rMhIuToQFG5nU
PxwXYLoVBkNVInFinBae54YvWbMlfmewN7m_3elajzilWKmns6EnQZbl3HZnJOiHmzs0
Connection : keep-alive
```

Lorsque le serveur reçoit une requête HTTP sans cookie, il en crée un et le transmet dans la réponse via la fonction *Set-Cookie*. Le navigateur reçoit la réponse et crée le cookie dans un fichier texte sur le disque dur de l'ordinateur.

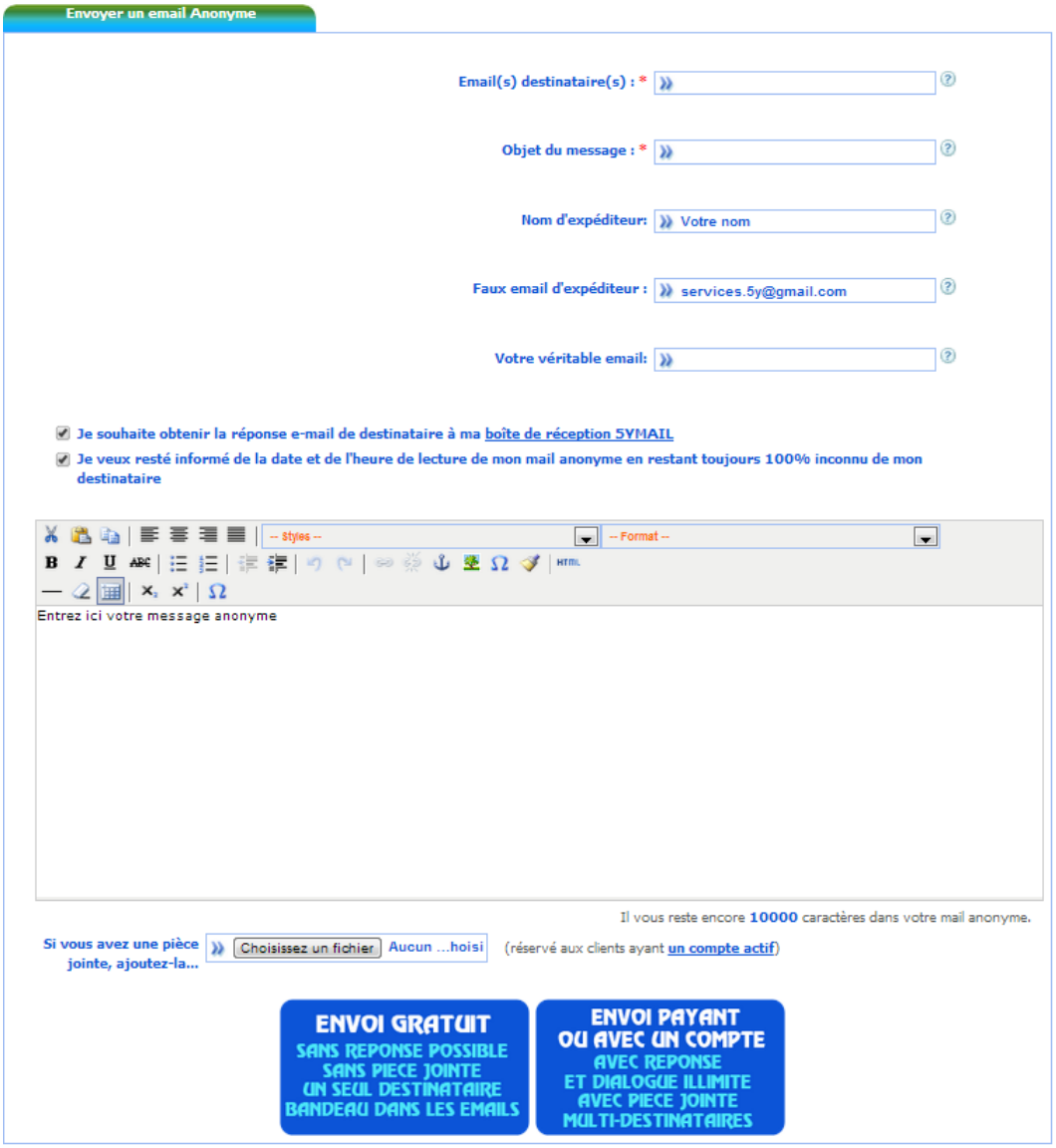
Lors de la connexion suivante, le navigateur fournit d'emblée le cookie retrouvé sur le disque,

pour le domaine ".google.be". Celui-ci contient exactement les mêmes informations que celles qui ont été fournies par le serveur, sans les méta-datas, composées du domaine auquel le cookie s'applique et de la date d'expiration.

Google explique les types de cookies qu'il utilise ainsi que leur utilisation comme "personnaliser les annonces" (http ://www.google.com/policies/technologies/types/).

C Exemple de site Internet offrant une boîte de messagerie de type pseudonymous

Le site "http ://email-anonyme.5ymail.com/" est un serveur de messagerie anonyme de type pseudonymous permettant l'envoi de mails anonymes vis-à-vis du destinataire.



D Exemple d'envoi de message à un réacheminеur de type I (Cypherpunk)

Première étape : obtenir la clé publique du réacheminеur de messages, cette information peut généralement être obtenue en lui envoyant un message dont le sujet est "remailer-key".

Deuxième étape : importer la clé publique dans le programme PGP ou GPG.

Troisième étape : rédiger le message dans un éditeur de texte en respectant le modèle suivant.

```
: :
Request-Remailing-To : <Adresse de messagerie du destinataire>

##
Subject : <Sujet>
<Corps du message>
```

Quatrième étape : chiffrer le message à l'aide de PGP ou GPG.

Cinquième étape : envoyer le message chiffré au réacheminеur en utilisant le modèle suivant.

```
: :
Encrypted : PGP

-----BEGIN PGP MESSAGE-----
<Message chiffré>
-----END PGP MESSAGE-----
```